



Caractérisation de toxines peptidiques par spectrométrie de masse à haute résolution.

Loïc Quinton

► To cite this version:

Loïc Quinton. Caractérisation de toxines peptidiques par spectrométrie de masse à haute résolution.. Chimie. Ecole Polytechnique X, 2006. Français. NNT : . pastel-00002201

HAL Id: pastel-00002201

<https://pastel.archives-ouvertes.fr/pastel-00002201>

Submitted on 29 Jul 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



THESE

Présentée pour obtenir le grade de

DOCTEUR DE L'ÉCOLE POLYTECHNIQUE

Spécialité : Chimie

Par

Loïc QUINTON

Caractérisation de toxines peptidiques par spectrométrie de
masse à haute résolution

Directrice de thèse : Julia Chamot-Rooke

Soutenue le 28 septembre 2006 devant la commission d'Examen composée de :

M. Gérard Bolbach

Président

M. Edwin De Pauw

Rapporteur

M. Pierre Escoubas

Rapporteur

Mme Muriel Delepierre

Examineur

Mme Odile Schiltz

Examineur

Mme Julia Chamot-Rooke

Directrice de Thèse

Remerciements

Ces travaux de thèse ont été réalisés sous la direction de **Julia Chamot-Rooke**. Je tiens tout particulièrement à la remercier de sa présence, de son soutien et de ses conseils. Je te remercie également de m'avoir permis, au cours du temps, de faire évoluer notre relation professionnelle en relation amicale. Merci pour ces trois années plus qu'enrichissantes.

Un grand merci également à **Gilles Ohanessian**, directeur du laboratoire des mécanismes réactionnels, pour m'avoir accueilli au sein de son laboratoire et ainsi permis de m'intégrer à la grande famille de la spectrométrie de masse.

J'adresse mes plus sincères remerciements à Messieurs **Edwin De Pauw** et **Pierre Escoubas** d'avoir accepté d'être les rapporteurs de ce travail. Merci aussi à Mesdames **Muriel Delepierre** et **Odile Schiltz** et Monsieur **Gérard Bolbach** d'avoir accepté de participer à l'évaluation de ce travail.

Merci également à tous les chercheurs du CEA avec qui j'ai eu la chance de collaborer. Je remercie plus particulièrement **Nicolas Gilles** et **Frédéric Ducancel** qui m'ont fait partager leur passion des toxines et qui ont toujours répondu à mes innombrables questions avec la plus grande des patiences. Merci également à **Denis Servent** et à **Jordi Molgo** pour leurs participations constructives dans les différents projets.

Je tiens à remercier spécialement **Jean-Pierre Le Caër** de m'avoir enrichi professionnellement mais aussi et surtout pour la complicité et l'amitié que nous avons tissées autour de la protéomique.

Merci également à **Thorsten**, mon collègue et ami allemand, pour tous ces moments plus agréables les uns que les autres. Le programme de nos journées situé entre échanges scientifiques, détente et (fous) rires, resteront gravés dans ma mémoire.

Merci enfin à **tout le personnel du DCMR** pour ces trois années mémorables que j'ai passées au laboratoire. Merci à Guillaume, Michel, Christian, Anne-Pascale,...

Merci à tous...

Je n'oublierai pas de remercier également,

Delphine,

Je ne te remercierai jamais assez de tout ce que tu fais pour moi. Merci pour ton soutien et ta complicité. Merci de partager mon quotidien et de me remonter le moral quand il va mal. Merci de partager mes joies, mes peines. Merci de t'être intéressée à mon travail et d'avoir cherché avec moi des solutions ... En fait, merci pour tout ton amour.

Mes parents et mes sœurs,

Merci de m'avoir donné le goût et les moyens de faire des études. Merci de m'avoir soutenu durant ma longue formation. Merci de vous être tenus au courant de l'avancée de mes recherches. Merci de votre réconfort. Merci pour tout. Cette réussite est aussi la vôtre.

Ma famille,

Merci à l'ensemble de ma (nombreuse) famille pour leur soutien mais aussi pour les discussions animées des repas familiaux et pour les questions récurrentes : « Mais quand est-ce que tu auras réellement fini ? », « A quoi servent tes recherches ? » et les « C'est pas dangereux de travailler avec des venins? »...

Un grand merci à vous tous,

Table des abréviations

ADN : Acide Désoxyribonucléique

ADNc : Acide Désoxyribonucléique Complémentaire

AP : Affinité protonique

ARN : Acide Ribonucléique

ARNm : Acide Ribonucléique Messenger

BIRD : Blackbody Infrared Radiation Dissociation

CAD : Collisional Activated Dissociation

CID : Collisional Induced Dissociation

CRM : Charge Residue Model

Da : Dalton (unité de mesure de masse moléculaire, équivalente au grammes/mole)

ECD : Electron Capture Dissociation

eV : electron volt ($1\text{eV}=1,602.10^{-19}\text{ J}$)

ESI : Electrospray Ionization

FT-ICR : Fourier Transform - Ion Cyclotron Resonance

IEM : Ion evaporation mechanism (mécanisme d'évaporation d'ions)

ISD : In Source Decay

IRMPD : Infrared Multiphoton Dissociation

LC : Liquid Chromatography (chromatographie liquide)

MALDI : Matrix-Assisted Laser Desorption Ionization (désorption ionisation laser assistée par matrice à pression atmosphérique)

MS : Mass Spectrometry (spectrométrie de masse)

MS/MS : Mass Spectrometry/Mass Spectrometry (Spectrométrie de masse en tandem)

m/z : rapport masse sur nombre des charges

nanoESI : nano-electrospray ionization

nanoLC : Liquid nano-chromatography

pH : potentiel Hydrogène

ppm : partie par million

Q-TOF : Quadrupole-Time of Flight

RCPG : Récepteur couplé à la protéine G

RF : radiofréquence

RMN : Résonance Magnétique Nucléaire

RT-PCR : Reverse Transcriptase – Polymerase Chain Reaction

RX : Diffraction des rayons X

SDS-PAGE : Sodium Dodecylsulfate Polyacrylamide Gel Electrophoresis

SORI-CAD: Sustained Off-Resonance Irradiation – Collision Activated Dissociation

SRTX : Sarafotoxine

TCEP : Tris(carboxyethyl)phosphine

Introduction

De nombreux organismes vivants, procaryotes ou eucaryotes, sont capables d'élaborer des substances toxiques. En biologie animale, on appelle *venins* des poisons d'origine animale susceptibles d'être utilisés comme arme d'attaque ou de défense envers un animal d'une autre espèce ou envers l'homme. Ces venins, substances complexes formées par sécrétion dans des glandes spécialisées, sont dans la plupart des cas soit injectés, soit projetés sur un prédateur potentiel ou une proie en vue de la paralyser ou de la tuer. On peut en extraire des *toxines* (du grec *toxicon* : poison pour flèche), qui sont des molécules (chimiques ou biologiques) bien définies ayant un effet physiologique nocif plus ou moins spécifique, ainsi que d'autres composés.

L'action d'un venin sur une proie doit être extrêmement rapide. Tout concourt en effet soit à une paralysie fulgurante, indispensable à la nutrition soit un effet dissuasif envers un agresseur. Les toxines agissent en général à très faible dose, et sont souvent des poisons très violents. Elles peuvent être classées selon leur mode d'action ou selon leur composition chimique mais d'autres classifications ont été proposées, notamment en fonction de l'animal qui les produit (cobratoxine de cobra par exemple).

C'est aux chimistes du XIX^{ème} siècle et en particulier à Lucien Bonaparte, frère de Napoléon, que l'on doit les toutes premières études rationnelles sur les venins. Des précipitations successives à l'alcool et à l'éther lui permettent d'isoler une fraction d'un venin de vipère *Vipera berus*, qu'il appelle « vipérine ». Fin observateur, Lucien Bonaparte note que cette fraction possède des propriétés biologiques rappelant celles des ferments digestifs.

En 1896, Calmette publie un petit volume présentant les techniques d'immunisation contre les venins de serpent. Cette brochure contribuera grandement au développement du traitement moderne des morsures de serpents : l'injection d'antivenins.

Il faut ensuite attendre la fin des années 60 et l'application des techniques séparatives aux biomolécules pour que les questions posées par F. Fontana au XVIII^{ème} siècle sur le mode d'action des venins et la nature de leur contenu, puissent trouver un début de réponse. Plus récemment encore, l'explosion de l'analyse protéomique et le développement de techniques de séparation et d'analyse toujours plus performantes et plus sensibles ont permis l'émergence d'une nouvelle discipline : la vénomique (étude du génome et du protéome des venins). En effet, les principaux composés des venins sont des protéines ou des peptides (plus de 90% du poids de venin séché), et c'est à ces composés que la plupart des propriétés toxiques et biologiques des venins sont dues.

Aujourd'hui, les venins représentent, de par la grande variété de leurs constituants, une source très riche de composés biologiquement actifs. Il est maintenant établi que les toxines qu'ils contiennent ont des activités variées et représentent des cibles de choix pour la recherche médicale, pharmacologique et agronomique. Comprendre le mode d'action de ces toxines et en synthétiser des analogues passe par une détermination précise de leur structure chimique, de leur architecture et des éléments qui assurent leur spécificité avec la contrainte que la faible quantité de matériel souvent disponible rend l'ensemble des processus de séparation et de caractérisation délicats. Le développement récent de techniques de spectrométrie de masse particulièrement bien adaptées à l'étude de composés biologiques thermolabiles (electrospray, nano-electrospray ou MALDI) permet désormais de caractériser des produits naturels présents à l'état de trace, ce qui n'était pas envisageable par les techniques d'analyse biochimiques classiques. Cependant, si l'utilisation en synergie d'un ensemble de techniques (résonance magnétique nucléaire, spectrométrie de masse, dégradation d'Edman) permet de déterminer précisément la plupart des structures des toxines peptidiques - ou non peptidiques - extraites des venins, l'élucidation de certaines toxines originales (du point de vue de leur structure et de leur activité) reste épineuse. Le développement de méthodes d'analyse alternatives donnant accès à l'ensemble des toxines biologiquement actives est donc d'un grand intérêt.

C'est dans ce contexte que se situe ce travail de thèse dont l'objectif était de *mettre au point de nouvelles stratégies d'analyse de toxines par spectrométrie de masse à haute résolution*, et plus particulièrement par FT-ICR (Résonance Cyclotronique Ionique à Transformée de Fourier).

Au cours de ce travail, nous nous sommes intéressés : d'une part à des toxines de cônes marins (mollusques utilisant un harpon relié à un appareil venimeux pour atteindre et paralyser leur proie) et d'autres part à des toxines de serpent. Plusieurs axes de recherche différents ont été explorés.

Tout d'abord la recherche de peptides vasoconstricteurs analogues des endothélines humaines (sarafotoxines) dans le venin brut du serpent *Atractaspis irregularis* a été réalisée par une approche basée à la fois sur la spectrométrie de masse FT-ICR et le clonage moléculaire. Ce travail fait l'objet du premier chapitre de cette thèse.

La caractérisation de nouvelles conotoxines dans les venins de *Conus virgo* et de *Conus ermineus* (chapitre 2) avec et sans couplage entre le spectromètre de masse FT-ICR et une chaîne de nano-chromatographie liquide a permis de comparer les avantages et limitations

de ces deux stratégies différentes. Ces venins de cônes ont été examinés en détail dans le cadre d'une recherche de ligands spécifiques pour des récepteurs du système nerveux central.

Enfin le troisième et dernier chapitre est consacré au séquençage *de novo* d'une toxine de haut poids moléculaire (8 kDa) et de ses homologues marqués d'un fluorophore, par les approches « bottom-up » (digestion de la protéine et fragmentation des peptides obtenus) et « top-down » (fragmentation de la protéine entière en phase gazeuse). Cette dernière approche innovante nous a permis d'évaluer l'intérêt de la technique de dissociation par capture d'électron (ECD) pour le séquençage de toxines. La toxine séquencée est spécifique des sous-récepteurs α_{1A} -adrénergiques et constitue donc un outil pharmacologique unique.

Ces trois chapitres sont précédés d'une introduction générale, à caractère bibliographique, faisant le point sur les techniques de spectrométrie de masse (particulièrement FT-ICR) utilisées pour l'analyse de biomolécules et l'état de l'art concernant la « vénomique » des cônes et de serpents. L'annexe regroupe les parties instrumentale et expérimentale ainsi que les principes des techniques biochimiques employées.

Une des perspectives de ce travail est de proposer au biologiste de nouveaux outils d'analyse donnant accès à des structures peptidiques parfois très complexes en s'affranchissant d'étapes de purification longues et fastidieuses. La connaissance de la structure peut alors conduire à une compréhension de la fonction à l'échelle moléculaire.

Introduction générale

Spectrométrie de masse et venins animaux

I. La spectrométrie de masse

A. Aspects Généraux et Historique

La spectrométrie de masse, en tant que technique d'analyse, a effectué un bond prodigieux depuis une décennie, comme en témoigne l'attribution du prix Nobel de chimie 2002. L'aspect le plus spectaculaire est sans doute la mise au point de nouvelles sources d'ionisation, qui permettent aujourd'hui d'analyser des échantillons de tout type (des petites molécules organiques aux macromolécules biologiques) et sous diverses formes (solide, liquide ou gazeux).

La nécessité de volatiliser l'échantillon, par chauffage sous vide, avant de pouvoir l'ioniser, a longtemps cantonné la spectrométrie de masse à l'étude de composés volatils (ou volatilisables, après dérivation chimique par exemple), excluant l'analyse de tous les composés de haut poids moléculaire et thermolabiles. Il faut attendre la fin des années 1980, et l'émergence de nouvelles techniques d'ionisation dites « douces » comme l'électrospray (ESI) et l'ionisation-désorption laser assistée par matrice (MALDI), pour que la spectrométrie de masse mette le pied dans le domaine de la biologie. Aujourd'hui, la spectrométrie de masse est devenue un outil incontournable en analyse protéomique (identification de protéines, séquençage de peptides, caractérisation de modifications post-traductionnelles), et trouve des applications dans des domaines aussi variés que la chimie supramoléculaire ou l'imagerie de tissus biologiques. Sa sensibilité et son couplage facile avec des méthodes de séparation (chromatographie gazeuse ou liquide) en font également une méthode de choix pour le contrôle qualité et le dosage de diverses molécules présentes à l'état de traces dans des mélanges complexes (contrôle anti-dopage par exemple).

Un spectromètre de masse peut être schématisé comme un assemblage de trois blocs bien distincts : une source d'ions, un analyseur et un détecteur (*Figure 1*).

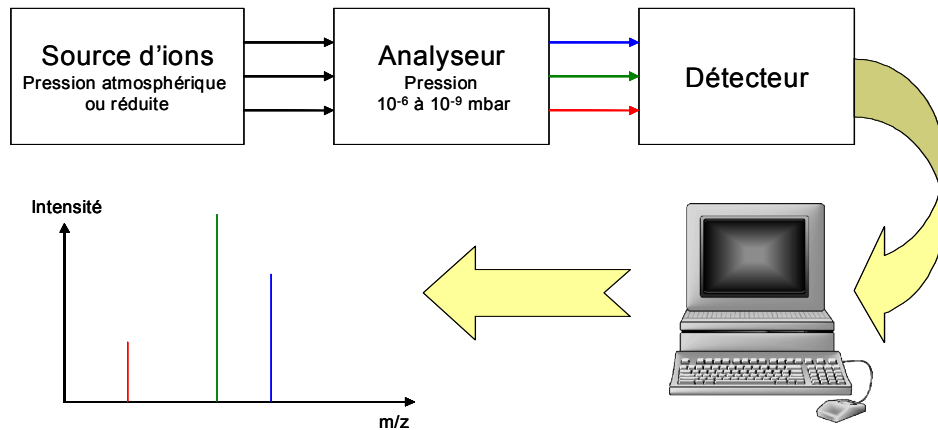


Figure 1: Schéma d'un spectromètre de masse, de la formation des ions à l'obtention d'un spectre

- La source d'ionisation permet d'ioniser le composé à analyser et de le transférer en phase gazeuse. Différents types de sources existent en fonction principalement du type d'échantillon à analyser (composé organique, biologique...). Les sources d'ionisation les plus employées pour les composés organiques sont les sources par Ionisation Electronique ou Ionisation Chimique alors que celles utilisées couramment pour les biomolécules sont l'electrospray (ESI) ou le nano-electrospray (nanoESI) ainsi que l'ionisation par désorption laser assistée par matrice (MALDI). Les ions formés dans la source sont ensuite conduits dans l'analyseur par l'intermédiaire de lentilles électrostatiques.

- L'analyseur est situé dans une enceinte à très basse pression (10^{-6} à 10^{-9} mbar) assurée par un système de pompage progressif (pompe à palettes primaires puis pompes turbomoléculaires). Le rôle de l'analyseur est de trier les ions selon leur rapport masse/ nombre de charge, noté m/z . Ce tri se fait de manière différente selon l'analyseur utilisé. Par exemple, dans un analyseur de type "temps de vol" (TOF) les ions seront séparés en fonction du temps qu'ils mettent pour traverser le tube de vol, alors que dans une cellule ICR, ils seront séparés en fonction de leur fréquence de rotation dans le champ magnétique. Au fur et à mesure que les ions sont triés, ils atteignent sélectivement le détecteur du spectromètre de masse.

- Le détecteur est chargé de "classer" les ions séparés précédemment. En général, les détecteurs jouent également le rôle de premier amplificateur de signal comme c'est le cas pour le channeltron ou les gallettes de micro canaux. Le signal est ensuite amplifié et traité par un système informatique qui, au final, nous donne le résultat sous la forme d'un spectre de masse, c'est-à-dire une représentation graphique du type *intensité du signal* = $f(m/z)$.

D'un point de vue historique, la spectrométrie de masse trouve ses racines à l'université de Cambridge, dans les études physiques de J.J. Thomson^{1,2} en 1910 sur la conductivité électrique des gaz. Thomson reçoit le prix Nobel en 1906 pour sa découverte de l'électron en 1897. Il est considéré comme le premier inventeur du spectromètre de masse, alors appelé spectrographe. Les ions formés étaient séparés par la différence de courbure de leur trajectoire sous l'influence d'un champ électrique et la détection s'effectuait par collision sur des plaques de type photographique. En 1909, Sir J.J. Thomson invite F.W. Aston à l'assister dans son travail sur la séparation des isotopes du néon. La première guerre mondiale retarde leurs travaux mais en 1919, F.W. Aston achève son amélioration du spectrographe en focalisant en vitesse les ions émergeant par un champ électromagnétique. Cette focalisation lui permet d'observer la séparation d'isotopes de masses peu différentes (1 unité de masse) dont les isotopes 20 et 22 du néon. Après plusieurs mesures, il découvre 212 isotopes d'éléments naturels^{3,4} et obtient le prix Nobel de Chimie en 1922. En parallèle, A.J. Dempster, de l'université de Chicago améliore en 1918 la résolution de son spectromètre de masse en focalisant les ions en direction⁵. Il met au point la première source à impact électronique en 1922, source qui permet d'ioniser un mélange de composés en phase gazeuse en le faisant interagir avec un faisceau d'électrons. La découverte des derniers isotopes du palladium, du platine, de l'or et de l'iridium en 1935 met un terme à l'utilisation de la spectrométrie de masse pour ce type de recherche, à tel point que F.W. Aston se montre alors très sceptique quant à l'avenir de la spectrométrie de masse !

La seconde guerre mondiale joue un rôle non négligeable dans la relance de la technique. En 1939, les tensions politiques ne cessant de s'intensifier en Europe, les physiciens américains décident de diriger leurs recherches vers la création d'une arme nucléaire basée sur la fission d'un noyau d'uranium. En 1940, dans le Minnesota, le jeune physicien Alfred Nier construit un spectromètre de masse capable de séparer les isotopes ^{235}U et ^{238}U , et confirme la théorie selon laquelle seul l'isotope 235 peut conduire à une fission alors que l'isotope 238, lui, reste stable. L'industrie chimique, et particulièrement l'industrie pétrolière, joue également un rôle important dans l'essor de la spectrométrie de masse en l'appliquant à l'analyse de petits composés organiques. En 1942, la *Consolidated Engineering Corporation* fabrique le premier spectromètre de masse commercial pour l'*Atlantic Refinery Company*, mettant en évidence l'attrait des industries pétrolières pour cette technique d'analyse. La spectrométrie de masse n'a alors cessé d'évoluer et d'innover afin de répondre aux demandes toujours plus nombreuses des chimistes (qui voulaient mesurer des masses toujours de plus en plus élevées) puis des biologistes. Les différents analyseurs développés sont :

- *L'analyseur par temps de vol* : en 1948, A.E.Cameron découvre le principe de l'analyse par temps de vol des ions (TOF) mais l'analyseur TOF ne sera vraiment décrit qu'en 1955 par W.C. Wiley et I.H. McLaren⁶. Il est ensuite considérablement amélioré en 1972 par V.I. Karatev qui introduit le mode reflectron qui permis d'obtenir des spectromètres TOF bien plus résolutifs.

- *Les analyseurs quadripolaires* : en 1953, W. Paul et H. Steinwedel brevètent ce type d'analyseur qui englobe les quadripôles et les pièges à ions⁷. Cinq ans plus tard, en 1958, W. Paul et ses collègues de l'Université de Bonn décrivent le premier spectromètre à quadripôle dans *Zeitschrift für Physik*. Il se voit décerner le prix Nobel de Physique en 1989.

- *L'analyseur à résonance cyclotronique ionique* : Dès 1951, H. Sommer applique la résonance cyclotronique à la spectrométrie de masse (ICR). Les ions sont soumis à un champ électrique radiofréquence et à un champ magnétique uniforme. Les ions oscillent alors à leur fréquence de résonance cyclotronique, inversement proportionnelle à leur rapport m/z . Cependant, ce type d'analyseur connaît un essor considérable, lorsqu'en 1974, M.B. Comisarov et A.G. Marshall⁸ appliquent la transformée de Fourier à l'analyse des spectres obtenus par ICR (*voir la partie sur la FT-ICR*).

- *L'Orbitrap* : Il s'agit d'une technologie très récente développée par A. Makarov⁹. L'orbitrap repose sur le piégeage des ions par un champ électrostatique quadripolaire qui crée l'oscillation des ions piégés. Ces oscillations, proportionnelles au rapport $(m/z)^{-1/2}$, sont détectées par la mesure du courant induit et traduites en spectre de masse par la transformée de Fourier.

Parallèlement à l'émergence de ces analyseurs, d'autres sources d'ions que l'impact électronique de Dempster (1922), se développent. Et c'est ainsi que notamment l'ionisation chimique¹⁰ (CI, 1966, M.S.B. Munson et F.H. Field), l'émission ionique secondaire par bombardement d'une surface par un faisceau d'ions primaires¹¹ (SIMS, 1971, A. Benninghoven et E. Loebach), la désorption plasma¹² (PDMS, 1976, R.D. Mac Farlane et D.F. Torgeson) et l'émission ionique secondaire en présence de matrice¹³ (FAB, M. Barber, 1981) apparaissent. C'est grâce à ces méthodes et notamment au SIMS et au FAB que les premières mesures de masses de molécules de type peptides et sucres peuvent être effectuées. Cependant, la forte concentration d'échantillon nécessaire ainsi que les présences de nombreux ions métastables résultant de la fragmentation des ions moléculaires formés ne permirent pas à la spectrométrie de masse de s'imposer dans l'analyse de biomolécules. Il faut attendre la fin des années 80 et l'émergence des méthodes d'ionisation douces grâce aux travaux des équipes d'Hillenkamp et Karas^{14,15}, et de Tanaka¹⁶ sur l'ionisation par MALDI et

ceux de John Fenn¹⁷ sur l'ionisation par électrospray pour que la spectrométrie de masse s'impose enfin comme une méthode de choix en biologie structurale. Ces deux dernières techniques d'ionisation et particulièrement l'électrospray largement utilisé durant cette thèse seront décrites plus en détail dans le paragraphe suivant.

Toutes ces découvertes liées à la spectrométrie de masse ont permis à cinq scientifiques d'obtenir un prix Nobel (*Figure 2*).

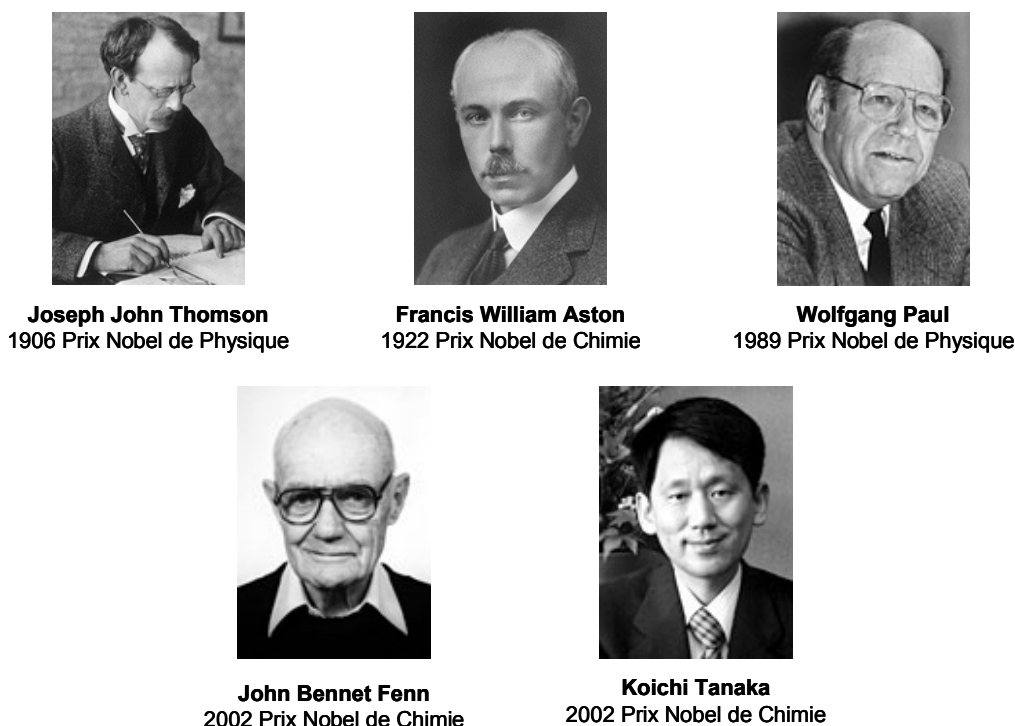


Figure 2: Les cinq prix Nobel de la spectrométrie de masse

B. Les sources d'ionisation adaptées aux biomolécules

B.1. L'electrospray

L'électronébulisation (ou electrospray) est une technique qui a été mise au point par J.B. Fenn sur la base de travaux plus anciens de R.L. Dole¹⁸ et J. Zeleny¹⁹. Il s'agit d'une technique d'ionisation à pression atmosphérique qui permet de former des ions en phase gazeuse par des mécanismes de désolvatation peu énergétiques n'induisant quasiment pas de fragmentation. C'est la raison pour laquelle l'electrospray est considérée comme une technique d'ionisation douce par opposition à d'autres techniques comme l'impact

électronique, technique très énergétique qui conduit à un grand nombre de fragmentations au moment de l'étape d'ionisation.

Les mécanismes de l'ionisation par électronébulisation sont, aujourd'hui encore, assez mal définis et sujets à controverse. Trois étapes principales permettent la formation d'ions en phase gazeuse : (i) la formation des gouttelettes chargées à l'extrémité d'un capillaire métallisé, (ii) l'évaporation du solvant et les explosions coulombiennes, et finalement (iii) l'obtention d'ions désolvatés en phase gazeuse à partir des gouttelettes fortement chargées issues des deux premières étapes.

○ **La formation de gouttelettes chargées à l'extrémité d'un capillaire métallisé :** l'échantillon est introduit dans un capillaire métallique (ou une aiguille métallisée) porté à un fort potentiel électrique de plusieurs kilovolts (kV). La grande différence de potentiel créée entre ce capillaire et une contre électrode, placée quelques millimètres plus loin, induit une séparation des charges à l'intérieur du liquide par un phénomène électrophorétique de migration des électrolytes (cations et anions). Lorsque le capillaire représente l'électrode positive, les espèces cationiques vont alors s'accumuler à son extrémité tandis que les anions vont observer un déplacement inverse en s'accumulant à l'intérieur, jusqu'à s'oxyder à la surface du capillaire (*Figure 3*).

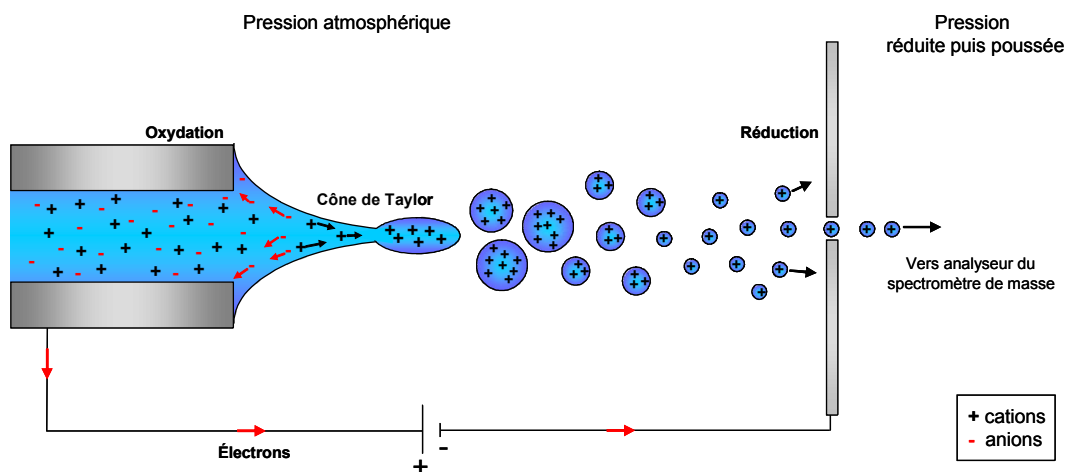


Figure 3: Schéma simplifié du fonctionnement d'une source électrospray (d'après P. Kebarle)²⁰

Les cations vont être attirés par l'électrode négative (contre-électrode) et les anions par l'électrode positive (capillaire). La séparation des charges va créer une réelle déformation du liquide à l'extrémité du capillaire, déformation qui met en jeu un équilibre entre les forces électrostatiques et la tension de surface du liquide. Lorsque les forces électrostatiques atteignent la valeur de la tension de surface du liquide, la déformation est conique et porte le

nom de cône de Taylor²¹. Une faible augmentation du champ électrique suffit alors pour déstabiliser le cône qui se rompt à son extrémité, dispersant la solution en nombreuses gouttelettes multichargées. Après cette étape, les gouttelettes se déplacent sous l'influence du champ électrique intense mais également sous celle du gradient de pression créé entre la source et l'entrée du spectromètre de masse. C'est au cours de ce trajet qu'a lieu l'évaporation du solvant des gouttelettes.

○ **Processus d'évaporation du solvant et explosions coulombiennes.** Dans l'espace entre l'extrémité du capillaire et la contre-électrode, les gouttelettes chargées issues du cône de Taylor vont entrer en collision avec le gaz résiduel et subir une évaporation du solvant alors que le nombre de charges reste constant. L'évaporation du solvant est favorisée par la température souvent élevée de la source d'ionisation (80-140 °C). La diminution du rayon R des gouttelettes possédant une charge constante q conduit à l'augmentation de la répulsion électrostatique des charge à sa surface jusqu'à ce que les gouttelettes atteignent la limite de stabilité de Rayleigh donnée par :

$$q_{Ry} = 8\pi(\epsilon_0\gamma R^3)^{1/2}$$

avec ϵ_0 permittivité du vide, R le rayon de densité de charges et γ la tension de surface stabilisant la goutte

Cette équation donne les conditions pour lesquelles les forces de répulsions électrostatiques deviennent égales à celles de tension de surface. A partir de la limite de Rayleigh, la gouttelette chargée devient instable et on observe une explosion coulombienne (fission des gouttelettes). Comme cela été abordé succinctement par Rayleigh²² en 1886 puis confirmé par des études ultérieures de Gomez et Tang²³, la fission des gouttelettes s'effectue par émission d'un jet très fin de gouttelettes (*Figure 4*), sans réelle relation de symétrie.

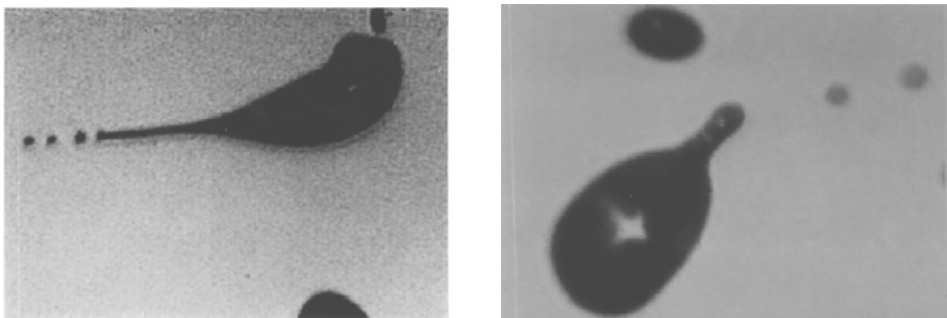


Figure 4 : Photographie de la fissions des gouttelettes chargées suite à l'évaporation du solvant d'ionisation, d'après Gomez et Tang²³.

Ce phénomène se reproduit jusqu'à ce que les nouvelles gouttelettes aient à nouveau atteint la limite de Rayleigh.

○ **Formation d'ions désolvatés en phase gazeuse à partir de petites gouttelettes fortement chargées** : deux mécanismes sont proposés pour expliquer ce phénomène. Le premier, le modèle des résidus chargés ou « charged residue model » (CRM) est décrit par Dole¹⁸ en 1968. Son hypothèse est que la succession d'explosions coulombiennes peut conduire à des gouttelettes de plus en plus petites jusqu'à ce que la dernière « gouttelette » ne possède plus qu'un seul et unique ion (*Figure 5a*). L'évaporation totale du solvant résiduel permet de ce fait l'obtention d'ions multichargés en phase gazeuse.

En 1979, une seconde théorie, le modèle de l'évaporation ionique ou « *ion evaporation model* » (IEM) basée sur de nouveaux mécanismes est proposée par Iribarne et Thomson²⁴. Leur théorie repose, comme celle de Dole, sur une succession initiale de fissions de type coulombiennes qui réduit la taille des gouttelettes en augmentant leur densité de charge. La différence avec la théorie de Dole repose sur le fait qu'Iribarne et Thomson proposent que lorsque les gouttelettes atteignent un rayon très petit (de l'ordre de 10 nm) et une densité de charges très élevée, le champ électrostatique à leur surface est assez intense pour expulser directement des ions en phase gazeuse (*Figure 5b*). Ce phénomène, appelé évaporation ionique, suggère que l'explosion coulombienne ne constitue pas le mécanisme majoritaire pour les petites gouttelettes de rayon $R \leq 10$ nm.

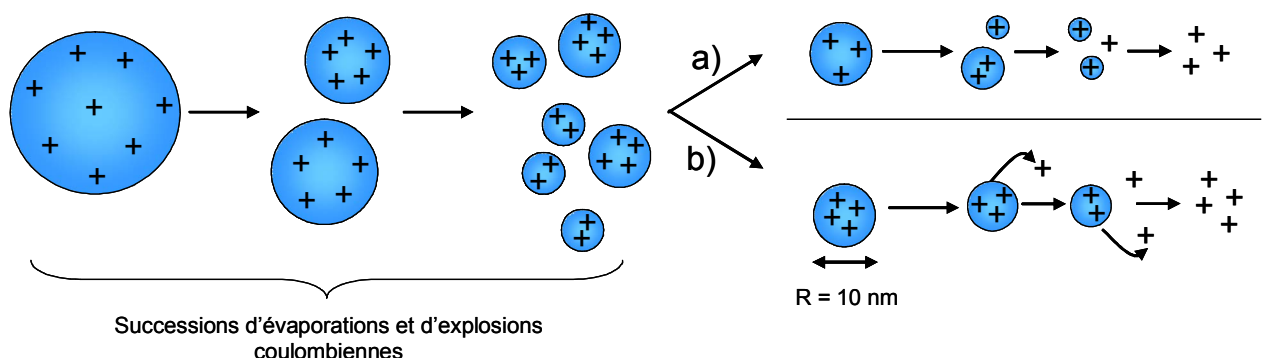


Figure 5: Les deux théories de l'obtention d'ions désolvatés en phase gazeuse à partir de gouttelettes chargées – (a) Modèle des résidus chargés par Dole – (b) Modèle de l'évaporation ionique par Iribarne et Thomson

Aucune de ces méthodes n'est aujourd'hui encore considérée comme la théorie unique qui permettrait de modéliser le passage de gouttelettes de solvant contenant des analytes à des ions en phase gazeuse²⁵. Cependant, il semble actuellement établi que dans le cas des grosses molécules fortement chargées, les ions en phase gazeuse soient formés par le mécanisme de résidus chargés de Dole^{26,27,28}. John B. Fenn émet l'hypothèse que cette théorie s'applique aux molécules de haut poids moléculaire à partir du moment où les dimensions linéaires de ces molécules sont sensiblement plus grandes que la gouttelette qui les contient²⁹.

Quel que soit le mécanisme responsable de la formation des ions en phase gazeuse, l'électrospray est aujourd'hui une méthode d'ionisation extrêmement utilisée, dans des domaines variés (biologie, chimie, pharmacie...).

B.2. Une miniaturisation de l'électrospray, le nano-electrospray

La miniaturisation de l'électrospray, dans sa version nano-electrospray (nanoESI), a été développée par Matthias Wilm et Matthias Mann^{30,31} à l'EMBL en Allemagne au milieu des années 90. Le principe est le suivant : un faible volume d'échantillon (2-3 μL) est introduit dans une aiguille métallisée dont l'orifice de sortie possède un diamètre de l'ordre de 1 à 2 μm . Une différence de potentiel est alors appliquée entre cette aiguille et la contre électrode placée quelques mm plus loin dans le spectromètre de masse afin de créer le spray, dont le débit peut être évalué à quelques dizaines de nanolitres par minute, typiquement 20-100 $\text{nL}\cdot\text{min}^{-1}$. Ce débit étant très faible, aucun gaz de nébulisation n'est nécessaire pour former le spray, et les tensions appliquées sont beaucoup plus faibles (quelques centaines de volts contre plusieurs milliers de volts en electrospray). On peut également noter que la température d'une source nano-electrospray est inférieure à celle d'une source electrospray car la quantité de solvant à évaporer dans les gouttelettes est moindre.

Le nanoESI présente de nombreux avantages par rapport à l'ESI. Tout d'abord, cette technique est moins consommatrice d'échantillon puisque quelques microlitres d'une solution à une concentration de l'ordre du μM (ou moins) permettent d'enregistrer un spectre. Cet aspect est important lorsque les échantillons étudiés sont disponibles en faible quantité, ce qui est généralement le cas pour les études protéomiques.

Un second intérêt des sources nanoESI est qu'elles se révèlent plus sensibles que l'ESI. Wilm et Mann évaluent le gain de sensibilité de leur source nanoESI à 510 fois ce qu'ils obtenaient avec leur source electrospray initiale³¹. Bien que cette valeur soit difficilement généralisable, il est admis que le nanoESI permet d'obtenir une sensibilité au moins 100 fois supérieure à

celle de l'ESI³¹. La principale explication à ce phénomène est que le faible diamètre de sortie de l'aiguille nanoESI permet de former des gouttelettes initiales très petites qui vont conduire rapidement à la formation d'ions en phase gazeuse puisque le volume de solvant à évaporer est très faible. Ceci a pour conséquence d'améliorer l'efficacité d'ionisation (rapport charge/volume supérieur à celui de l'ESI).

Un autre avantage du nanoESI est sa tolérance aux sels présents dans l'échantillon^{31,32}: des études ont ainsi montré que l'on peut analyser, en nano-electrospray, des échantillons jusqu'à dix fois plus chargés en sels que ne le permet l'utilisation d'une source ESI conventionnelle³².

Néanmoins, il y a quelques inconvénients liés à cette technique principalement liés à l'aiguille utilisée pour réaliser le spray. Ces aiguilles sont généralement en silice et sont recouvertes d'un métal conducteur. Le fait que les aiguilles soient fabriquées par étirement du verre jusqu'à l'obtention d'une extrémité très petite (1 à 2 µm de diamètre), conduit dans de nombreux cas à son obturation. Il faut donc "ouvrir" mécaniquement ces aiguilles afin de pouvoir initier le spray. Cette ouverture n'étant absolument pas reproductible, il est impossible de contrôler avec précision le diamètre de sortie du capillaire une fois ouvert. Or, comme ce diamètre joue un rôle important sur la taille initiale des gouttelettes, qui elle-même va avoir une influence sur le spectre obtenu (état de charge des ions, formation d'adduits salins, intensités des différentes espèces dans le spectre), la reproductibilité des résultats obtenus par nanoESI n'est pas bonne.³³.

Malgré cet inconvénient, le gain de sensibilité apporté par l'utilisation d'une source nano-electrospray en fait une technique d'ionisation de choix pour l'analyse de biomolécules (mesure de masse de protéines, séquençage *de novo* de peptides...).

B.3. Principes du MALDI

L'ionisation par désorption laser assistée par matrice (MALDI) a été décrite pour la première fois par Karas et Hillenkamp^{14,15} et par Tanaka¹⁶ dans les années 80.

Son principe repose sur l'utilisation d'une matrice absorbant dans l'ultraviolet, co-cristallisée avec les analytes, et déposée sur une plaque métallique. L'ionisation/désorption est induite par un faisceau de photons issu d'un laser pulsé émettant dans l'ultra violet (*Figure 6*). La majorité des lasers utilisés (N₂) émettent à une longueur d'onde de 337 nm.

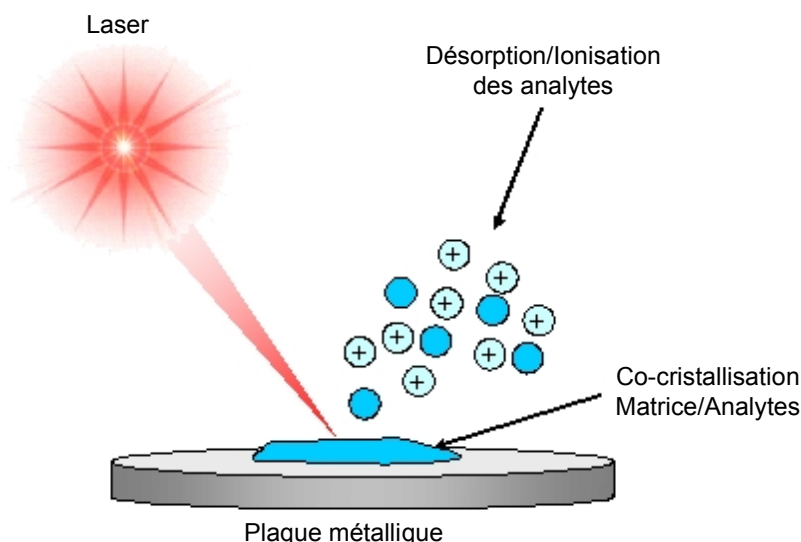


Figure 6: Principe de la désorption/ionisation MALDI

L'irradiation du mélange de matrice et d'échantillon conduit à l'ionisation de la matrice. On suppose que l'ionisation des molécules d'analyte est ensuite due au transfert d'un proton de la matrice ionisée à l'analyte probablement par réaction ion-molécule en phase gazeuse. L'énergie transmise au système par le pulse laser est absorbée par la matrice qui subit une expansion en phase gazeuse au cours de laquelle elle entraîne des molécules d'analytes.

Le choix de la matrice est primordial pour observer une bonne ionisation/désorption des composés. Généralement, cette matrice doit être soluble dans les solvants utilisés pour les analytes, elle doit être capable de co-cristalliser avec les analytes, elle ne doit pas se sublimer dans le vide de la source MALDI et elle doit absorber à la longueur d'onde du laser. Toutes ces contraintes font que les matrices sont, en général, des composés polaires aromatiques. Classiquement, l'acide dihydroxybenzoïque (DHB), l'acide α -cyano-4-hydroxycinnamique (α -CHCA), et l'acide 3,5-dimethoxy-4-hydroxycinnamique (SA) sont utilisés pour l'étude des peptides et des protéines.

Le MALDI est particulièrement bien adapté aux exigences de la protéomique et constitue de ce fait l'une des premières méthodes d'ionisation pour la caractérisation de composés biologiques. L'obtention majoritaire d'espèces mono-chargées permet de générer des spectres de masses simples et facilement interprétables. L'intérêt de cette technique repose également sur sa bonne tolérance aux sels et aux détergents, fréquemment rencontrés en analyse protéomique.

C. La spectrométrie de masse FT-ICR

Il existe différents types d'analyseurs dont les analyseurs de type quadripolaire avec les triple-quadripôle (tQ) et les pièges à ions (IT), les analyseurs à temps de vol (TOF), les analyseurs à secteurs magnétiques, les spectromètres de masse à résonance cyclotronique ionique et transformée de Fourier (FT-ICR) et les récents Orbitrap. Chacun d'entre eux possède des caractéristiques et des performances analytiques qui lui sont propres. D'une manière générale, ils peuvent être définis par ces différents paramètres :

1- **La gamme de masse** qui correspond aux limites minimales et maximale de m/z au-delà desquelles l'analyseur ne pourra plus mesurer de signal.

2- **La transmission**, qui est le rapport entre le nombre d'ions formés en source et celui qui arrive au détecteur.

3- **La sensibilité** qui correspond à la quantité d'échantillon nécessaire pour détecter un signal (interprétable) dans le spectre de masse.

4- **La vitesse de balayage** qui est le temps que va mettre l'analyseur à établir un spectre pour une gamme de masse donnée.

5- **La résolution**, qui représente la capacité de l'analyseur à séparer deux composés de m/z voisins. La définition de la résolution (à partir d'un seul pic) s'exprime par $R = M/\Delta M$ où M est la masse de l'ion d'intérêt et ΔM , la largeur du pic à mi-hauteur.

6- **La précision** sur la mesure de masse qui représente la faculté de mesurer une masse la plus proche possible de la masse théorique de la molécule analysée. Cette précision s'exprime le plus souvent en « partie par million » (ppm) calculée selon :

$$Erreur(ppm) = \frac{|M_{\text{exp}} - M_{\text{th}}| \cdot 10^6}{M_{\text{th}}}$$

avec M_{exp} représentant la masse mesurée expérimentalement et M_{th} la masse réelle de la molécule analysée.

Parmi les différents analyseurs, les spectromètres de masse FT-ICR tiennent une place particulière, du fait de leurs très grandes performances en terme de résolution et de précision sur la mesure de masse. Ce travail de thèse ayant été réalisé par spectrométrie de masse FT-ICR, seule cette technique sera détaillée.

C.1. La cellule ICR

Dans notre spectromètre de masse (Apex III Bruker 7T), les ions sont formés dans une source externe à la cellule de piégeage puis acheminés *via* différentes lentilles électrostatiques au centre de la cellule située au cœur d'un électroaimant supraconducteur de 7 Tesla (*Figure 7*).

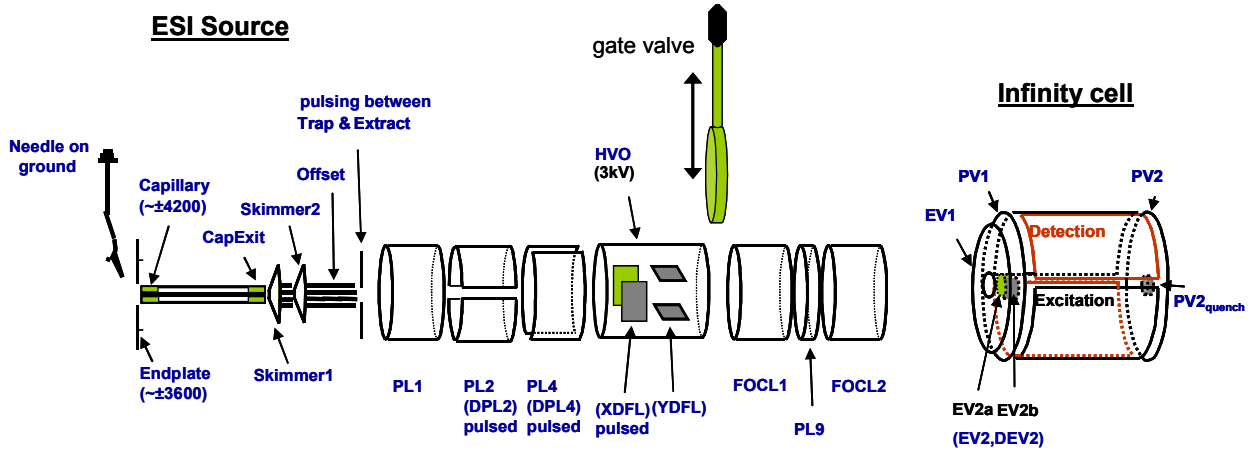


Figure 7: Schéma du FT-ICR APEX III Bruker, de la source électrospray à la cellule ICR Infinity Cell

Un multipôle de stockage (hexapôle pour l'APEX III) peut être inséré entre la source et la cellule afin d'accumuler les ions avant de les envoyer dans la cellule ICR. Cette dernière, placée dans un vide poussé (10^{-10} mbar environ), représente l'élément clé du spectromètre puisqu'elle va non seulement servir d'analyseur (piégeage des ions) mais aussi de détecteur (excitation et détection des ions).

Le principe de la spectrométrie de masse à résonance cyclotronique ionique repose sur le fait qu'un ion qui se déplace dans un champ magnétique uniforme B subit la force de Lorentz. Cette force le conduit à adopter une trajectoire hélicoïdale autour d'un axe parallèle à celui de B dont la projection dans le plan xy est un mouvement circulaire uniforme (*Figure 8*).

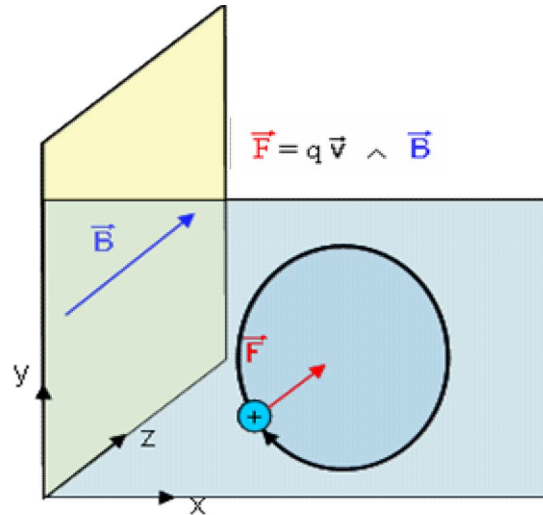


Figure 8: Mouvement dans le plan (x,y) d'une particule chargée dans un champ magnétique uniforme

Ce mouvement de rotation est appelé *mouvement cyclotron* et sa fréquence ν_c ne dépend que du champ magnétique B et du rapport masse/charge selon :

$$\nu_c = \frac{\omega}{2\pi} = \frac{qB}{m} = \frac{1.535611 \times 10^7 B}{m/z} \text{ Hz (m/z en Da)}$$

Par conséquent, la mesure de cette fréquence de rotation permet une détermination du rapport m/z de l'espèce considérée. Pour donner un ordre de grandeur, dans un champ magnétique de 7 T, cette fréquence vaut 1,1 MHz pour un ion de m/z 100, 215 kHz pour un ion de m/z 500 et 54 kHz pour un ion de m/z 2000.

C.2. Le piégeage des ions

Néanmoins, en l'absence de champ électrique, le mouvement hélicoïdal des ions autour de l'axe z conduirait à leur fuite hors de la cellule. Il est par conséquent nécessaire, pour les maintenir à l'intérieur du piège, d'appliquer un potentiel répulsif (typiquement 0,5 à 2V pour les ions positifs) sur les deux plaques parallèles au plan xy . Ces plaques sont appelées plaques de piégeage *Figure 9*.

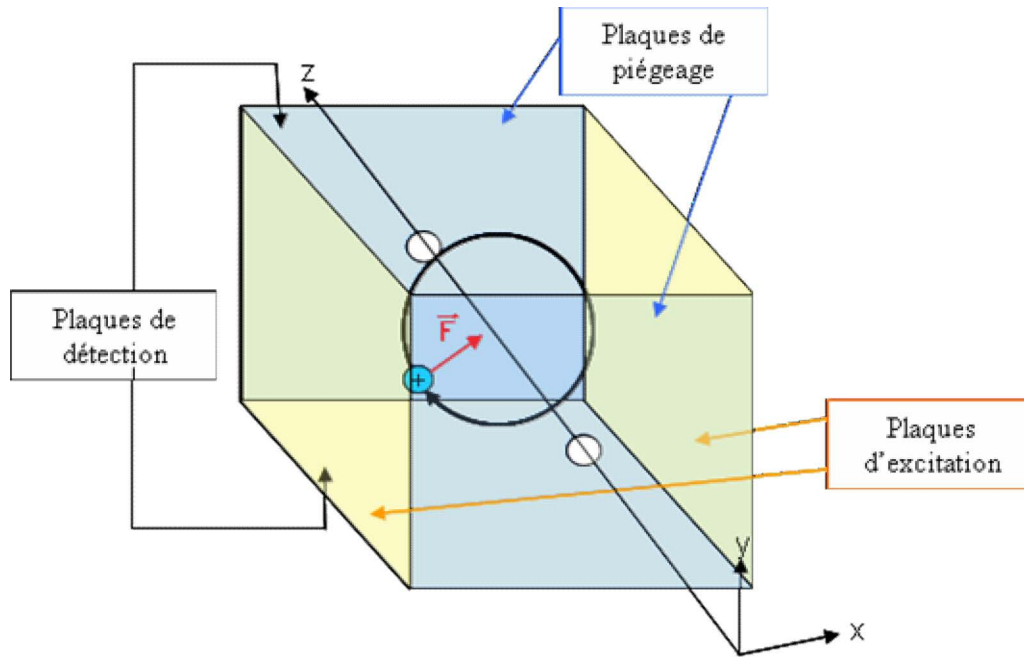


Figure 9: Schéma d'une cellule ICR (piège de Penning)

S'il était possible d'appliquer un champ électrique homogène et parallèle au champ magnétique par l'intermédiaire de plaques de longueurs infinies, le mouvement de l'ion serait hélicoïdal avec des oscillations périodiques suivant l'axe z . Cette fréquence d'oscillations ou fréquence de piégeage ν_t est principalement due aux répulsions successives des deux plaques de piégeage et peut être exprimée par :

$$\nu_t = \frac{1}{2\pi} \sqrt{\frac{2 \cdot q \cdot V_{trap} \cdot \alpha}{m \cdot a^2}}$$

avec V_{trap} le potentiel de piégeage, a la distance entre les deux plaques et α le facteur géométrique ($\alpha=2$ dans le cas idéal)

A titre d'exemple, pour un aimant 7T, un potentiel de piégeage de 1V dans une cellule dont les plaques sont espacées de 6 cm (conditions typiques), la fréquence de piégeage sera de 1650 Hz pour un ion de m/z 1000.

Cependant, il n'est pas possible de se situer dans le cas idéal d'un champ électrique homogène qui suppose des plaques infinies. Le champ électrique ne peut pas être exactement parallèle au champ magnétique et une composante radiale est alors à prendre en compte. Par conséquent, un troisième mouvement propre au mouvement de déplacement des charges piégées s'ajoute aux deux précédents mouvements (cyclotron et piégeage). Ce mouvement appelé mouvement

« magnétron » s'oppose à la force de Lorentz et se traduit par un abaissement de la fréquence cyclotronique (7) et (8).

$$\nu_+ = \frac{\nu_c}{2} + \sqrt{\left(\frac{\nu_c}{2}\right)^2 - \frac{\nu_t^2}{2}} \quad \text{Fréquence cyclotron réduite}$$

$$\nu_- = \frac{\nu_c}{2} - \sqrt{\left(\frac{\nu_c}{2}\right)^2 - \frac{\nu_t^2}{2}} \quad \text{Fréquence magnétron}$$

Aux basses masses $\nu_+ \approx \nu_c$ mais il existe des masses plus élevées (fréquences plus faibles), pour lesquels cette approximation n'est plus valable. Ce point sera atteint d'autant plus rapidement que le potentiel de piégeage est élevé, d'où l'intérêt de travailler avec des potentiels de piégeage faible (typiquement 0,5 à 2V pour les ions positifs) même si l'efficacité de piégeage en est amoindrie.

C.3. Obtention d'un spectre de masse

Une fois piégés, la mesure de la fréquence cyclotronique des ions n'est pas possible directement. Bien qu'ayant une fréquence de rotation identique, tous les ions de même rapport m/z ne présentent pas un mouvement cohérent. En effet, les ions ont été produits à des temps variables et avec des énergies cinétiques différentes, ce qui se traduit par une phase (position sur l'orbite) et un rayon variables. Afin de pouvoir mesurer les fréquences cyclotroniques, il est nécessaire que ce mouvement global devienne cohérent, c'est-à-dire que tous les ions de même rapport m/z se présentent sur la même phase et avec le même rayon. Ce résultat est obtenu par une procédure d'excitation-détection détaillée ci-dessous.

- Excitation des ions piégés

Une fois piégés, les ions de même rapport m/z ont la même fréquence cyclotronique, mais ne forment pas un paquet cohérent : la position de leur axe de rotation et la phase de leur mouvement ne sont pas égales. De plus, le rayon de leur orbite cyclotronique est trop faible pour induire un quelconque signal sur les plaques de détection. L'application d'un pulse de radiofréquence sur les *plaques d'excitation* (Figure 9) à leur fréquence cyclotron, permet de rendre le mouvement des ions cohérent et de les conduire sur un même rayon, proche des plaques de détection. En effectuant un balayage de fréquence, il est possible d'exciter des populations d'ions de m/z différents qui adoptent alors une même trajectoire circulaire avec

des fréquences de rotation différentes. Cette étape permet non seulement d'obtenir des paquets d'ions décrivant un mouvement cohérent mais aussi de les rapprocher des deux plaques de détection afin de pouvoir analyser ce mouvement. Dans le cas idéal, il conviendrait d'exciter les ions sur une orbite la plus grande possible afin d'augmenter le signal et la sensibilité obtenus. Cependant, ceci n'est pas applicable puisque l'inhomogénéité du champ électrique près des plaques conduirait à l'éjection des ions. C'est pourquoi il est usuel de ne pas exciter les ions sur des rayons plus grands que 0,8 fois le rayon de la cellule. Plusieurs types de champs électriques sinusoïdaux peuvent être appliqués pour exciter les ions. La *Figure 10* montre trois types d'excitations différentes :

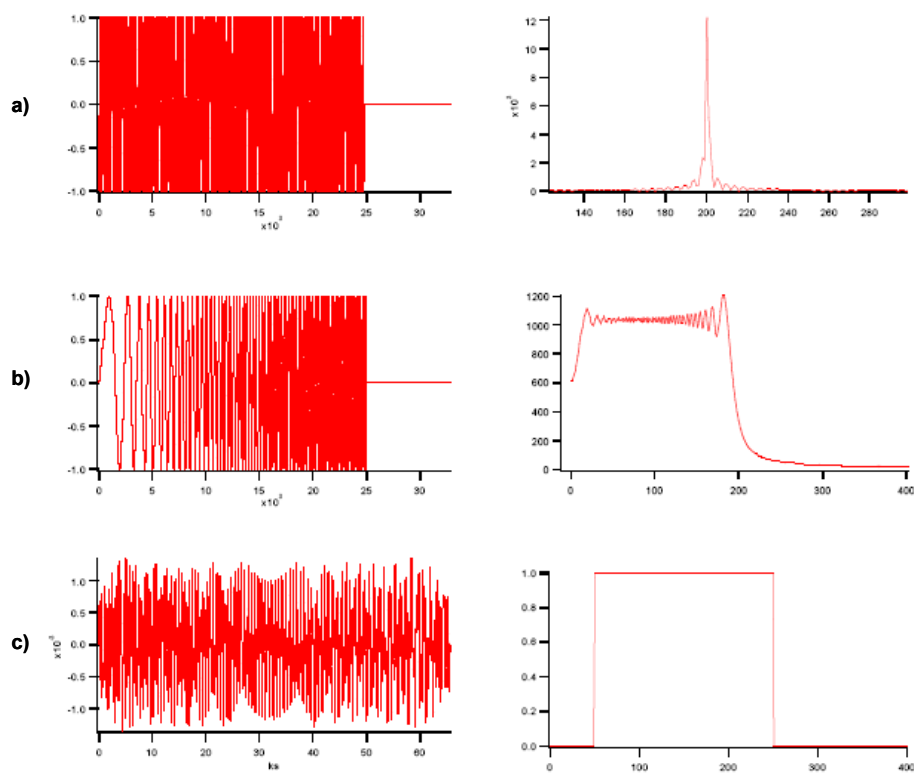


Figure 10: Types d'excitations usuellement utilisées, a gauche en échelle de temps et à droite leurs composantes obtenues après transformée de Fourier – (a) sinusoïde mono fréquence tronquée, (b) balayage de fréquence (*chirp*), (c) balayage par SWIFT

La première excitation permet d'exciter à une seule fréquence. Elle n'est donc pas vraiment utile dans le cas où il faut détecter tous les ions présents dans la cellule, cas le plus couramment rencontré. Il faut alors effectuer une excitation mono fréquence pour chacun des ions présents. A cela, il est préférable d'utiliser un mode d'excitation par balayage de fréquence (*Figure 10- b*) qui va balayer toute la gamme de fréquences des ions en une seule impulsion. Le balayage par SWIFT pour *stored-waveform inverse Fourier Transform* (*Figure*

10- c) permet de calculer l'onde d'excitation à partir du domaine de m/z étudié. Cependant, le SWIFT n'est employable que sur les instruments équipés d'un générateur de fréquences.

- Détection des ions

La rotation cohérente des ions sur une orbite large crée un courant induit sur les deux *plaques de détection* détectable par une électronique adaptée. Ce courant est proportionnel à la charge totale du paquet d'ions ayant un mouvement cohérent et ne dépend pas de la fréquence cyclotronique des ions.

La collision des ions avec des molécules de gaz résiduelles pendant le temps de mesure conduit à une perte de cohérence du paquet d'ion et donc une perte du signal. De ce fait, le signal obtenu correspond à une sinusoïde amortie exponentiellement. Or la précision en masse est directement liée à la précision sur la mesure de la fréquence de rotation cyclotron des ions piégés. Une mesure de masse précise nécessitera une pression la plus faible possible dans la cellule.

- Digitalisation et transformée de Fourier

Comme indiqué dans l'introduction, c'est la transformée de Fourier qui permet de mesurer en une seule expérience l'ensemble des fréquences de rotation des ions (*Figure 11*).

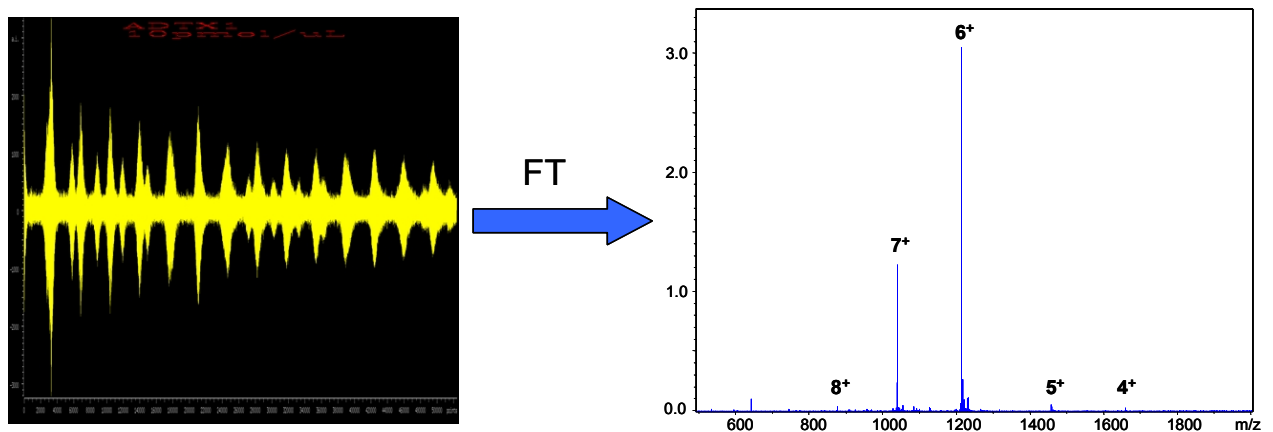


Figure 11 : Spectre de masse obtenu après Transformée de Fourier du signal

Le courant induit mesuré sur les plaques de détection correspond à un interférogramme (FID) où les sinusoïdes amorties de l'ensemble des ions présents dans la cellule sont additionnées les unes aux autres. La digitalisation de ce signal, suivie d'une transformation de Fourier permet de convertir cet interférogramme en un spectre contenant les

informations sur les fréquences (facilement convertibles en m/z) et les abondances des ions piégés dans la cellule. En additionnant plusieurs détectations successives, il est possible d'améliorer le rapport signal sur bruit du spectre transformé. Diverses méthodes de traitement numérique du signal sont également accessibles pour améliorer sa qualité. Nous pouvons citer notamment le *zero-filling* qui permet une interpolation entre les points dessinant le spectre (interpolation de Shannon) et l'*apodisation* qui permet l'affinage des pics présents sur le spectre.

C.4. Performances de la spectrométrie de masse FT-ICR

- Résolution et précision en masse

La particularité de la spectrométrie de masse FT-ICR qui est de ramener la mesure des rapports m/z à la mesure d'une fréquence, lui confère ses caractéristiques en termes de résolution. Dans des conditions adéquates, c'est-à-dire avec un vide très poussé dans la cellule permettant un enregistrement pendant un temps suffisamment long sans perte de signal, il est possible d'atteindre des résolutions supérieures à 10^6 pour un ion de m/z 1000 dans un champ magnétique de 7T. Cependant, ce type de mesure ne peut s'effectuer qu'en mode « hétérodyne », c'est à dire en se restreignant à une fenêtre de fréquences réduites. En mode large bande, il est possible d'atteindre une résolution de 200 000. Plus généralement, dans les analyses plus routinières où la résolution n'est pas optimisée, il est courant d'obtenir des résolutions de l'ordre de 60 000. La précision en masse dépend principalement de la résolution mais également d'autres paramètres comme la *charge d'espace*. En effet, la présence d'un très grand nombre d'ions piégés dans la cellule modifie le champ électrique et génère une force de répulsion qui va modifier la trajectoire des ions et ainsi modifier les mesures de masses. Ce phénomène pose un problème évident pour les étalonnages externes. En effet, pour que l'étalonnage externe soit le plus précis possible, il faut qu'il soit effectué avec une charge d'espace comparable à celle de l'échantillon, ce qui est difficilement contrôlable. Il est ainsi recommandé de travailler avec de faibles quantités d'ions pour limiter au maximum la déviation due à ce phénomène. On obtient alors une précision de l'ordre de quelques ppm. Cependant, l'étalonnage interne permet de s'affranchir de ce problème de charge d'espace puisque les ions calibrant sont présents dans la cellule en même temps que les ions de l'échantillon. On atteint alors couramment des précisions en masse de l'ordre de 1 ppm, voire moins.

- Sensibilité

Des calculs théoriques ont montré que le nombre minimum de charges ayant un mouvement cohérent, nécessaire pour détecter un signal était de l'ordre de 200. Ceci constitue une limite ultime de la sensibilité de ce type d'instrument. En pratique il faut tenir compte de l'efficacité de chacune des étapes permettant de passer de l'analyte neutre aux ions piégés et excités dans la cellule. Il faut également tenir compte des états de charge et de l'existence de massifs isotopiques qui augmentent le nombre de fréquences sur lesquelles se répartissent les ions. Les sensibilités *records* atteintes actuellement sur des peptides de masse 1000 Da sont de l'ordre de 10 attomoles consommées en nano-ESI. Dans un instrument commercial avec les protocoles standard, on peut obtenir un signal exploitable avec environ 100 femtomoles de produit consommé.

- Gamme dynamique

Un point sur lequel la FT-ICR n'est pas particulièrement bien placée est la gamme dynamique qui représente la capacité à discriminer des ions d'abondances différentes dans un même spectre. Le problème est lié à celui de la charge d'espace dans la mesure où les mouvements des ions de faible abondance sont perturbés par ceux des ions majoritairement présents dans la cellule. Il est admis qu'il est difficile de mesurer avec une bonne précision la masse d'un ion dont l'abondance est inférieure à 1% de celle de l'ion le plus abondant.

- Gamme de masse

La gamme de masse, quant à elle, est en théorie quasiment illimitée. La limite de masse inférieure dépend uniquement de l'électronique et plus précisément des capacités du digitaliseur et du générateur de fréquences. Pour un aimant de 7T, équipé d'un digitaliseur à 8 MHz, cela équivaut à une limite inférieure de m/z 29. En ce qui concerne la limite supérieure, elle dépend de la capacité de la cellule à piéger les ions. A cause de l'inhomogénéité radiale du champ électrique, il existe une fréquence pour laquelle on obtient $v_+ = v_-$. Au-delà de cette limite, la trajectoire des ions n'est plus stable dans la cellule. Dans le cas d'un aimant de 7T, cette limite est atteinte pour un ion de m/z 274.000. Malheureusement, des problèmes apparaissent avant cette limite, notamment à cause des limites sur les durées d'acquisition de signal. Les signaux des masses élevées correspondant à de basses fréquences se trouvent

considérablement élargis (m/z 274.000 équivaut à mesurer une fréquence de 393 Hz). De ce fait, la limite expérimentale couramment admise se situe à m/z 27.000 pour un aimant de 7 T.

C.5. Nouveaux spectromètres hybrides qFT-ICR et LIT-FTICR

Récemment des améliorations instrumentales ont été apportées aux FT-ICR avec l'apparition de spectromètres hybrides. Il s'agit notamment des spectromètres q-FTICR³⁴ (pour *quadripole-FTICR*) et LIT-FTICR³⁵ (pour *linear ion trap-FTICR*). La grande innovation de ces instruments est d'introduire un analyseur avant la cellule ICR (*Figure 12*).

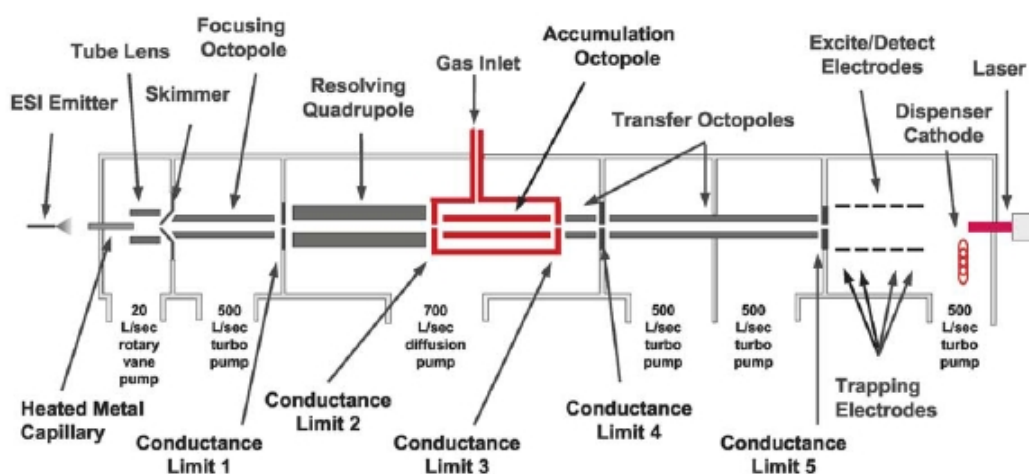


Figure 12: Représentation schématique d'un spectromètre de masse hybride quadrupole-FTICR³⁴

L'intérêt principal de ces nouveaux instruments est de pouvoir effectuer de la MS/MS rapide et efficace avant la cellule ICR. La sélection de l'ion parent s'effectue dans le premier analyseur (quadrupole ou piège linéaire) et les fragmentations dans une cellule de collision ou directement dans le piège linéaire. L'analyse des fragments obtenus s'effectue ensuite dans la cellule ICR. Ceci n'est pas possible dans les FT-ICR de la génération précédente pour lesquels l'isolation et la fragmentation sont réalisées dans la cellule et nécessitent une mise au point minutieuse des conditions expérimentales, rendant les analyses à haut débit de type LC-MS/MS inenvisageables. La sélection d'un ion donné avant son entrée dans la cellule permet également de le concentrer et améliore grandement la gamme dynamique. Par ailleurs, les instruments de type LIT-FT-ICR sont équipés d'un système permettant de contrôler le nombre d'ions rentrant dans la cellule, ce qui diminue les effets de charge d'espace et augmente la précision sur la mesure de masse (sub ppm).

C.6. Applications de la spectrométrie de masse FT-ICR

Les applications de la spectrométrie de masse FT-ICR sont variées. L'une d'entre elle est la mesure exacte de petits composés. Comme cela vient d'être décrit, un des avantages les plus puissants de la FT-ICR est sa capacité à déterminer des masses très précisément. Ainsi pour des ions monochargés < 500 Da, une composition élémentaire unique peut être proposée si la mesure de masse est réalisée avec une précision de l'ordre de 1 ppm³⁶. Il faut noter que le très haut pouvoir de résolution de la spectrométrie de masse FT-ICR permet des déterminations de masses exactes sur des mélanges très complexes pour lesquels plusieurs pics sont observés à la même masse nominale. Une calibration interne est généralement souhaitable pour ce type d'expérience. L'équipe d'Alan Marshall en Floride détient la plupart des records dans le domaine de l'analyse de mélanges complexes par FT-ICR, que ce soit sur des pétroles bruts³⁷ ou sur diverses sortes de vins³⁸ pour lesquels des milliers de constituants différents ont pu être déterminés sans séparation ni purification préalable. L'intérêt de pouvoir s'affranchir d'étapes de purifications ou de séparations longues et fastidieuses est évident pour tout ce qui relève de l'analyse protéomique et qui va conduire à des spectres très complexes : analyse différentielle par marquage isotopique³⁹, analyse du contenu protéique de cellules entières⁴⁰, séquençage top-down⁴¹...

D'autres applications, et notamment celles relatives à la protéomique seront abordées plus en détails dans le paragraphe de ce chapitre intitulé « *Apport de la spectrométrie de masse FT-ICR au séquençage de novo* ».

D. La spectrométrie de masse en tandem : application au séquençage de peptides et de protéines

D.1. Aspects généraux de la fragmentation

Les spectromètres de masse en "tandem" sont ceux qui permettent (i) dans une première étape, de sélectionner l'ion d'intérêt (appelé ion parent ou ion précurseur) (ii) dans une deuxième étape, d'analyser les ions fragments (ou ions fils) obtenus par la dissociation de l'ion parent. Selon les instruments utilisés, les étapes de sélection et d'analyse des fragments

vont pouvoir se faire soit dans l'espace (tQ, Q-TOF, TOF-TOF), soit dans le temps (Trappe ionique, FT-ICR, orbitrap).

Les sources d'ionisation utilisées pour l'analyse des biomolécules (ESI, nanoESI, MALDI) conduisant à des ions avec peu d'énergie interne, il faut induire la fragmentation par un apport d'énergie au système. Cet apport d'énergie peut se faire de différentes façons⁴² : collision avec un gaz inerte (fragmentations de haute ou basse énergie), collision avec une surface, irradiation par des photons, irradiation par des électrons (spécifique des FT-ICR).

La spectrométrie de masse en tandem permet d'obtenir des informations de structure de l'ion fragmenté (qu'il soit organique ou biologique) et permet, dans le cas de mélanges complexes, de sélectionner spécifiquement une espèce parmi les autres pour obtenir son spectre de fragmentation. Les différents types de fragmentation utilisables pour les biomolécules (et plus spécifiquement les peptides) sont détaillés dans le paragraphe suivant.

- Fragmentation avec un gaz inerte

Ce type de fragmentation est celui qui est le plus couramment rencontré et est utilisable sur tous les spectromètres de masse. On parle de dissociation induite par collision (CID) ou de dissociation activée par collision (CAD). Le principe est le suivant : l'ion d'intérêt est sélectionné puis accéléré dans une zone de haute pression où il va rentrer en collision avec un gaz chimiquement inerte. Dans le cas des FT-ICR, un (ou plusieurs) pulse de gaz est introduit dans la cellule ICR après sélection de l'ion. Dans une trappe ionique, c'est le gaz de thermalisation (He) qui sert de gaz de collision. Pour les autres instruments, la fragmentation a généralement lieu dans une cellule de collision, se trouvant sur la trajectoire des ions, préalablement remplie de gaz. Lors d'une collision entre un ion incident et une molécule neutre, le choc inélastique conduit à la transformation d'une partie de l'énergie cinétique de l'ion incident en énergie interne. L'énergie disponible au centre de masse est donnée par la formule :

$$E_{CM} = E_{lab} \frac{M_c}{M_c + M_i}$$

avec M_c la masse du gaz de collision, M_i la masse de l'ion, et E_{lab} son énergie cinétique

Cette équation met en évidence que plus le gaz de collision a une masse élevée (M_c), plus la valeur de l'énergie E_{CM} va tendre vers celle de l' E_{lab} et donc plus la collision va être efficace. A titre d'exemple, remplacer l'argon ($M_c=39,95$ uma) par du xénon ($M_c=131,29$ uma) permet

de gagner environ un facteur 3 en efficacité de fragmentation. C'est pourquoi, le xénon a été utilisé pour toutes les expériences de fragmentation par collisions menées dans cette thèse.

Le cas le plus favorable pour la fragmentation est $E_{CM}=E_{lab}$, atteint pour une valeur de M_c tendant vers l'infini. Ce cas limite est atteint dans la méthode de fragmentation induite par collision avec une surface (Surface Induced Dissociation ou SID) au cours de laquelle les ions incidents vont entrer en collision avec une surface métallique.

Cette équation permet également de comprendre que plus la masse de l'ion à fragmenter (M_i) est grande, plus l'énergie transférée au centre de masse est faible ce qui se traduit par une faible efficacité de la collision. Les fragmentations induites par collisions ne sont plus vraiment efficaces au-delà de 3000 unités de masse dans le cas d'un ion monochargé.

Comme rappelé plus haut, la dissociation induite par collision peut être séparée en deux catégories. Le CID de basse énergie pour des énergies de 1 à 100 eV (tQ, IT, FT-ICR) et le CID de haute énergie de l'ordre du keV (TOF-TOF et secteurs magnétiques)⁴².

- La fragmentation à basse énergie est liée au fait que le temps entre deux collisions est grand, ce qui permet aux ions de perdre une partie de leur énergie d'excitation. Afin d'obtenir une fragmentation, il est donc nécessaire de pouvoir faire plusieurs collisions successives afin d'augmenter progressivement l'énergie interne des ions. C'est ce que l'on appelle le régime multi-collisionnel. Le spectre obtenu est alors dominé par des produits formés par des mécanismes de faibles énergies d'activation et induites par la charge.

- La MS/MS de haute énergie est caractéristique des instruments à secteurs et également des TOF. L'énergie cinétique de l'ion incident est telle qu'une seule collision va conduire à la fragmentation *via* une excitation électronique. C'est un régime mono-collisionnel. L'énergie emmagasinée est rapidement convertie dans les différents modes de vibrations de l'ion, ce qui conduit à de multiples fragmentations. Le résultat est que le spectre obtenu contient un grand nombre d'ions fragments qui peuvent être obtenus par des mécanismes ayant de fortes barrières d'activation (remote fragmentation ou mécanisme à charge piégée).

- Fragmentation par collision avec une surface

Ce type de fragmentation monocollisionnelle, assez peu employée, permet de contrôler finement l'énergie donnée aux ions en faisant simplement varier E_{lab} . Cette technique ne nécessite pas l'emploi d'un gaz de collision et est parfois employée sur les FT-ICR⁴³.

- In Source Decay (ISD)

La dissociation en source ne fait pas partie des techniques de spectrométrie de masse tandem car elle n'implique pas de sélection de l'ion à fragmenter. La fragmentation est simplement réalisée en accélérant fortement l'ion parent dans une région de la source dans laquelle la pression est importante. Il en résulte de multiples fragmentations avec le gaz résiduel conduisant à des fragments.

- La capture d'électron dissociative (ECD)

Ce mode de dissociation consiste à irradier les ions multichargés par un faisceau d'électrons. La recombinaison entre un électron et l'ion conduit à la formation d'un cation radical et permet un gain d'énergie de l'ordre de quelques eV. Ce type de fragmentation, souvent associé à une technique d'activation douce de type IRMPD (Infra-Red Multi Photon Dissociation) est spécifique des FT-ICR et sera explicité un peu plus loin, dans la partie consacrée à la fragmentation de peptides par FT-ICR.

D.2. La fragmentation de peptides

D.2.a. Modèle du proton mobile

L'étude de la fragmentation des peptides est un thème de recherche développé depuis plusieurs dizaines d'années par différents groupes. L'objectif principal de ces études est non seulement d'être capable de prédire la fragmentation d'un peptide en fonction de sa séquence et par conséquent de pouvoir mettre au point des logiciels de séquençage beaucoup plus fiables que ceux qui existent actuellement, mais aussi d'utiliser ces connaissances pour diriger la fragmentation et l'utiliser judicieusement.

Pour expliquer la fragmentation d'un peptide portant un ou plusieurs protons en phase gazeuse, il existe un modèle général appelé « modèle du proton mobile »^{44,45}. Ce modèle propose que l'énergie apportée par la fragmentation soit suffisante pour permettre au proton, (dans la majorité des cas) de se détacher de son site de fixation initial pour aller échantillonner l'ensemble du peptide. Cet échantillonnage conduit à des formes "réactives" des peptides, *i.e.* des formes à partir desquelles la fragmentation est facile. A titre d'exemple, la protonation sur l'azote de la liaison amide, moins favorable thermodynamiquement que la protonation sur

l'oxygène, conduit à une élongation de la liaison amide et une rupture facile de cette liaison. Ainsi, une des hypothèses de ce modèle est que la fragmentation est dirigée par la charge (*charge-directed pathways*).

D.2.b. Sites de protonation

La protonation initiale du peptide a lieu sur le site le plus basique en phase gazeuse c'est-à-dire, par ordre décroissant de basicité, l'arginine, la lysine, l'histidine et l'extrémité N-terminale. La basicité très élevée de l'arginine fait que lorsque cet acide aminé est présent dans la séquence, le proton de la protonation est littéralement "séquestré" et n'a plus aucune possibilité de devenir mobile. Ainsi, en basse énergie, dans le cas de peptides monochargés comportant une arginine, il est courant d'observer des spectres de fragmentation comportant très peu de fragments.⁴⁶ W. Wisocki a montré le rôle de l'arginine sur la fragmentation en étudiant des peptides contenant cet acide aminé et des homologues sans arginine. Elle montre que dès que l'arginine est absente de la séquence, le proton redevient mobile et sa mobilisation conduit à de nombreuses fragmentations.

Ce modèle explique pourquoi il est important, dans le cas de peptides issus d'une digestion protéolytique à la trypsine *i.e.* où l'acide aminé C-terminal est soit une lysine soit une arginine, d'isoler et de fragmenter préférentiellement des peptides multichargés. La fragmentation des ions multichargés sera plus efficace que celle des ions monochargés du fait qu'au moins l'un des protons de protonation sera mobile. Ainsi, les instruments capables de réaliser des spectres MS/MS basse énergie en automatique (sélection automatique de l'ion parent suivie de sa fragmentation) en sortie de colonne nanoLC sont généralement programmés pour ne sélectionner que des ions multichargés.

Comme indiqué précédemment, le site de protonation initial n'est pas le site qui conduit à la fragmentation mais c'est la mobilité du proton dans la séquence et notamment son attachement à l'azote de la liaison peptidique qui mène à la rupture de cette liaison.

D.2.c. Nomenclature pour la fragmentation des peptides et mécanismes de formation des ions

La nomenclature adoptée pour la fragmentation des peptides est celle proposée par P. Roepstorff et J. Fohlman en 1984⁴⁷ (*Figure 13*).

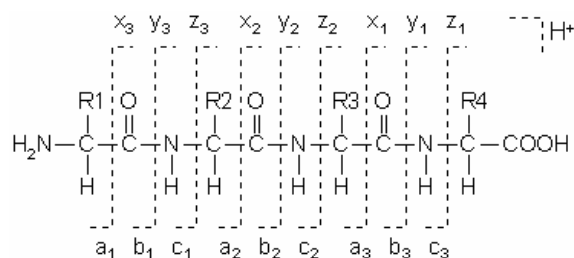
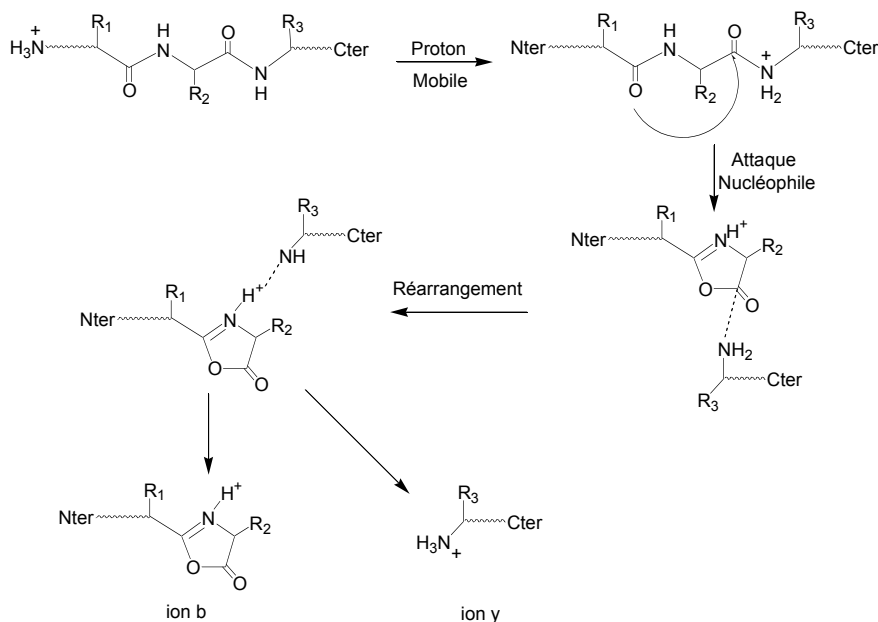


Figure 13: Nomenclature pour la fragmentation des peptides

Les ions sont notés a_i , b_i ou c_i s'ils comportent l'extrémité *N*-terminale et x_i , y_i ou z_i s'ils portent l'extrémité *C*-terminale. Le numéro indique le nombre d'acides aminés que contient l'ion. La rupture de la liaison peptidique conduit aux ions b et y . Une revue récente de B. Paizs et S. Suhai fait le point sur les mécanismes réactionnels (proposés sur la base de calculs de chimie quantique et d'expériences) permettant d'expliquer principalement la formation des ions b et y ⁴⁸. Ils montrent notamment que la formation de ces ions, par des collisions de basse énergie peut être expliquée par un mécanisme à plusieurs étapes (Figure 14).

Figure 14: Mécanisme de formation des ions y et b ⁴⁸

La première étape est le transfert du proton mobile situé sur l'extrémité *N*-terminale vers l'azote d'un amide d'une liaison peptidique. Cette étape est suivie d'une attaque nucléophile de l'oxygène du carbonyl de l'acide aminé du côté *N*-terminal sur le carbone du carbonyl de

l'acide aminé activé par la protonation. Cette attaque conduit, après réarrangement des différents constituants du complexe, à la formation d'une oxazolone protonée si le proton reste du côté *N*-terminal (ions *b*) et d'un peptide tronqué protoné si le proton reste du côté *C*-terminal (ions *y*). Bien que d'autres chemins de fragmentation existent pour former des ions *y* et *b* (attaques nucléophile des chaînes latérales, formation de dicétopipérazines ou d'aziridinones etc...), le mécanisme proposé ci-dessus est celui qui est applicable dans la majorité des cas.

Pour les spectres de fragmentation comportant plusieurs ions *y* ou *b* consécutifs, la séquence primaire devient facilement lisible puisque la différence de masse entre deux ions *y* ou deux ions *b* consécutifs correspond exactement à la masse d'un acide aminé (*Figure 15*).

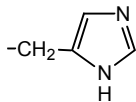
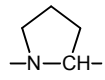
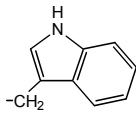
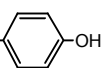
Nom	Code 1 lettre	Code 3 lettres	Masse monoisotopique (Da)	Acide aminé (NH-CHR-CO) R=
Alanine	A	Ala	71,0371	-CH ₃
Arginine	R	Arg	156,1011	-CH ₂ -(CH ₂) ₂ NHC(NH ₂)=NH
Asparagine	N	Asn	114,0429	-CH ₂ -CONH ₂
Aspartate	D	Asp	115,0269	-CH ₂ -COOH
Cystéine	C	Cys	103,0092	-CH ₂ -SH
Glutamate	E	Glu	129,0426	-CH ₂ -CH ₂ -COOH
Glutamine	Q	Gln	128,0586	-CH ₂ -CH ₂ -CONH ₂
Glycine	G	Gly	57,0215	-H
Histidine	H	His	137,0589	
Isoleucine	I	Ile	113,0841	-CH(CH ₃)-CH ₂ -CH ₃
Leucine	L	Leu	113,0841	-CH ₂ -CH(CH ₃) ₂
Lysine	K	Lys	128,0950	-CH ₂ -(CH ₂) ₃ -NH ₂
Méthionine	M	Met	131,0405	-CH ₂ -CH ₂ -S-CH ₃
Phénylalanine	F	Phe	147,0684	-CH ₂ -Ph
Proline	P	Pro	97,0528	
Sérine	S	Ser	87,0320	-CH ₂ -OH
Thréonine	T	Thr	101,0477	-CH(OH)CH ₃
Tryptophane	W	Trp	186,0793	
Tyrosine	Y	Tyr	163,0633	-CH ₂ - 
Valine	V	Val	99,0684	-CH(CH ₃) ₂

Figure 15: Table des 20 acides aminés naturels les plus courants

Les ions immoniums (*Figure 16*) représentent un autre type d'ions très intéressant d'un point de vue analytique. Contrairement aux séries *y* et *b*, les immoniums ne donnent pas d'information sur l'enchaînement des acides aminés dans la séquence primaire du peptide mais permettent d'obtenir des informations sur la nature des acides aminés présents dans la séquence, quelle que soit leur position.

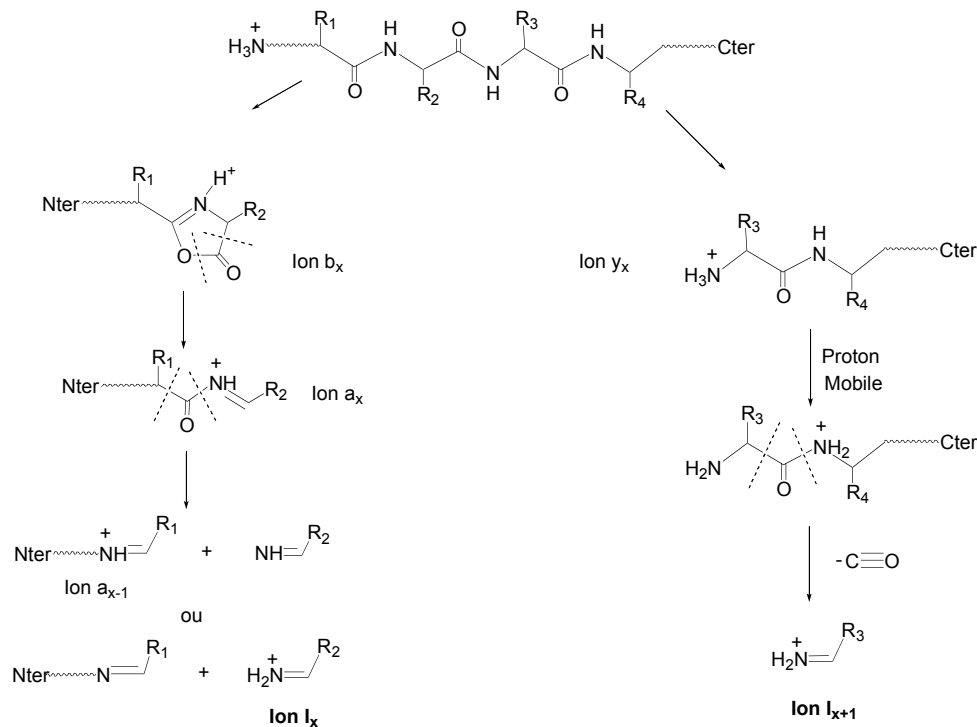


Figure 16: Mécanismes de formation des ions immoniums⁴⁸

Les ions immoniums sont issus de plusieurs ruptures dans le squelette polypeptidique. Notons également qu'en collision de faible énergie, il est possible d'observer des ions de type *a*. Ils découlent principalement de la perte d'une molécule de monoxyde de carbone par les ions *b* bien que d'autres mécanismes permettent également de les former. D'autres fragmentations, résultant de la perte de molécules neutres ($-\text{NH}_3$ ou $-\text{H}_2\text{O}$), sont couramment observées et dépendent de la nature des acides aminés présents dans le peptide. Enfin, des fragmentations internes peuvent aussi être présentes en faibles intensités dans les spectres. Les autres types d'ions ne sont pas (ou peu) observés en MS/MS de faible énergie.

Des ions caractéristiques de la chaîne latérale sont en revanche observés dans les spectres de fragmentation à haute énergie (ions *w* et *v*) et peuvent se révéler utiles pour distinguer les isomères leucine / isoleucine par exemple. Les spectres de fragmentation des peptides à haute

énergie sont très complexes car on peut en fait observer à peu près tous les types d'ions fragments et notamment l'apparition de cations radicaux.

Si la fragmentation par collision permet d'obtenir des informations très intéressantes sur la structure primaire d'un peptide, voire dans certains cas reconstruire totalement sa séquence, elle n'est pas particulièrement bien adaptée pour les peptides portant des modifications post-traductionnelles labiles (glycosylations, phosphorylations...). En effet, dans ces cas particuliers, la liaison la plus faible devient celle entre la modification et le squelette peptidique. De ce fait, c'est souvent la première liaison qui se rompt lorsque l'on apporte de l'énergie au système, rendant très difficile la localisation de la modification dans la séquence. Des techniques émergentes comme l'ECD pour les FT-ICR ou l'ETD pour les pièges ioniques permettent de contourner ce problème.

D.3. Apport de la spectrométrie de masse FT-ICR au séquençage *de novo*

L'utilisation de la spectrométrie de masse pour le séquençage *de novo* remonte à une trentaine d'années⁴⁹. Les performances de la spectrométrie de masse FT-ICR en font un outil tout à fait intéressant pour le séquençage de peptides voire même de protéines. Plusieurs types de fragmentation différents peuvent être envisagés sur un FT-ICR dont les plus utilisés sont : la fragmentation par collision avec un gaz inerte (SORI-CAD) et la dissociation par capture d'électron (ECD). Nous allons revenir en détail sur ces techniques.

D.3.a. Fragmentation par SORI-CAD

La dissociation activée par collisions et par excitation hors résonance (SORI-CAD ou Sustained Off Resonance Irradiation – Collision Activated Disociation)⁵⁰ est une technique d'activation couramment utilisée sur les instruments de type FT-ICR. Il s'agit d'une technique de dissociation de faible énergie (<200 eV) qui s'apparente à du CID « classique » en augmentant progressivement l'énergie interne des ions par collisions successives avec un gaz inerte jusqu'à ce que cette énergie soit suffisante pour induire une fragmentation. En SORI-CAD, les ions précurseurs sont excités à une fréquence d'irradiation légèrement décalée (500 Hz environ) par rapport à leur fréquence cyclotronique. Dans ce cas, les ions parents observent une succession d'accélérations/décélérations ce qui se traduit par des cycles d'augmentation et de diminution du rayon de leur trajectoire au sein de la cellule. En présence de gaz, ces mouvements conduisent à une augmentation de l'énergie interne jusqu'à

dissociation. L'énergie transmise aux ions par cette activation est plus faible que celle par activation résonante, moins utilisée en FT-ICR. Cet inconvénient est largement atténué par le fait qu'en SORI-CAD, il est possible d'utiliser des temps d'activation plus longs sans perte d'ions, augmentant le nombre de collisions observées.

Typiquement, un ou plusieurs pulse(s) de gaz sont introduits dans la cellule ICR avant l'excitation. Dans notre cas, le gaz de collision initial (argon) a été remplacé par un gaz plus lourd (le xénon) afin d'améliorer l'efficacité de la fragmentation. Dans le cas des peptides, le processus de SORI-CAD va rompre préférentiellement la liaison peptidique (liaison de plus faible énergie) conduisant majoritairement aux ions *y* et *b*.

L'intérêt de réaliser la fragmentation d'un peptide par collision avec un gaz inerte sur un instrument FT-ICR est lié principalement à la précision en masse qu'apporte ce type d'instrument, qui peut être très utile pour identifier précisément un acide aminé ou un groupe d'acides aminés (*tag*) dans la séquence.

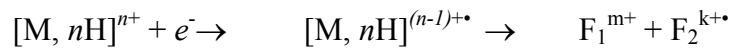
Excepté les isomères leucine et isoleucine, les 20 acides aminés communs ont des compositions élémentaires différentes, permettant de les différencier. Malgré cela, une faible précision sur la mesure de masse peut conduire à des ambiguïtés au niveau de la différenciation de certains acides aminés comme la glutamine (128,0586 Da) ou la lysine (128,0950 Da). A ces problèmes s'ajoute le fait que la combinaison de deux acides aminés peut avoir la même composition élémentaire qu'un autre acide aminé. Ceci est particulièrement problématique dans l'étude des spectres présentant peu de fragments. C'est le cas notamment du tag Gly-Gly qui présente exactement la même composition élémentaire que Asn (C₄H₆N₂O₂, 114,04292 Da).

La grande précision en masse qu'apporte la spectrométrie de masse FT-ICR permet de lever un très grand nombre d'ambiguïtés. Bien évidemment, elle permet de distinguer très facilement les acides aminés séparés de 1 Da (Asp/Asn et Glu/Gln) même pour des masses très élevées (masses de protéines entières). Mais surtout, les capacités de la FTMS font qu'elle peut différencier des écarts de masse bien plus faibles comme celui séparant la glutamine de la lysine ($\Delta M = 26$ mDa) ou le tag Val-Val de Pro-Thr ($\Delta M = 36$ mDa) ou de Arg-Gly ($\Delta M = 11$ mDa). D'un point de vue des modifications post-traductionnelles, elle peut même, dans certains cas, permettre de distinguer une sulfatation d'une phosphorylation ($\Delta M = 9,5$ mDa)^{51,52}. Ces distinctions seront d'autant plus faciles que la précision est excellente, ce qui est le cas lorsqu'il est possible de travailler en calibration interne.

D.3.b. La capture d'électron dissociative (ECD)

- Principes de l'ECD

La capture d'électron dissociative (ECD pour *Electron Capture Dissociation*) est une méthode de MS/MS assez récente permettant d'accéder à des types de fragmentations différents de ceux observées par les méthodes de MS/MS par collisions avec un gaz inerte. Les premières expériences d'ECD ont été menées par l'équipe de F. Mc Lafferty et particulièrement par R. Zubarev en 1998⁵³. Ils ont notamment observé qu'une protéine multiprotonée irradiée par un faisceau d'électrons thermiques (ou de faible énergie) était capable de capturer un électron pour former un cation radical conduisant à des ions fragments de type *c* et *z* selon :



D'un point de vue technique, les électrons sont formés par une cathode chauffée indirectement remplaçant le filament de source interne placé du côté opposé à la source externe de l'ICR, afin d'obtenir un flux suffisant pour que la capture d'électron soit efficace. Cette technique de dissociation se distingue particulièrement des autres techniques puisqu'elle ne conduit pas au clivage de la liaison peptidique mais à celui de la liaison N-C_α, du peptide en laissant intactes toutes les chaînes latérales. Ainsi, les modifications post-traductionnelles labiles (phosphorylation, ⁵⁴ O-glycosylation⁵⁵, N-glycosylation,...) restent attachées au squelette ce qui permet leur localisation et leur identification.

Un autre avantage de cette technique est que la fragmentation dépend assez peu de la séquence du peptide fragmenté puisque ce n'est pas un proton mobile qui induit la fragmentation en s'échantillonnant sur le squelette peptidique. En revanche, un point important, du point de vue des applications, est que l'ECD ne permet pas d'observer la rupture de la liaison Xxx-Pro en raison de la structure cyclique de la proline.

Le mécanisme initialement proposé était : la capture de l'électron par un site protoné de type ammonium quaternaire pour conduire à la formation d'un radical H[•] qui peut alors migrer dans le peptide vers les sites de haute affinité⁵⁶ (pont disulfures, carbonyles) et conduire à la formation d'ions fragments de type *c* et *z* par rupture de la liaison N-C_α. La *Figure 17* résume ces étapes.

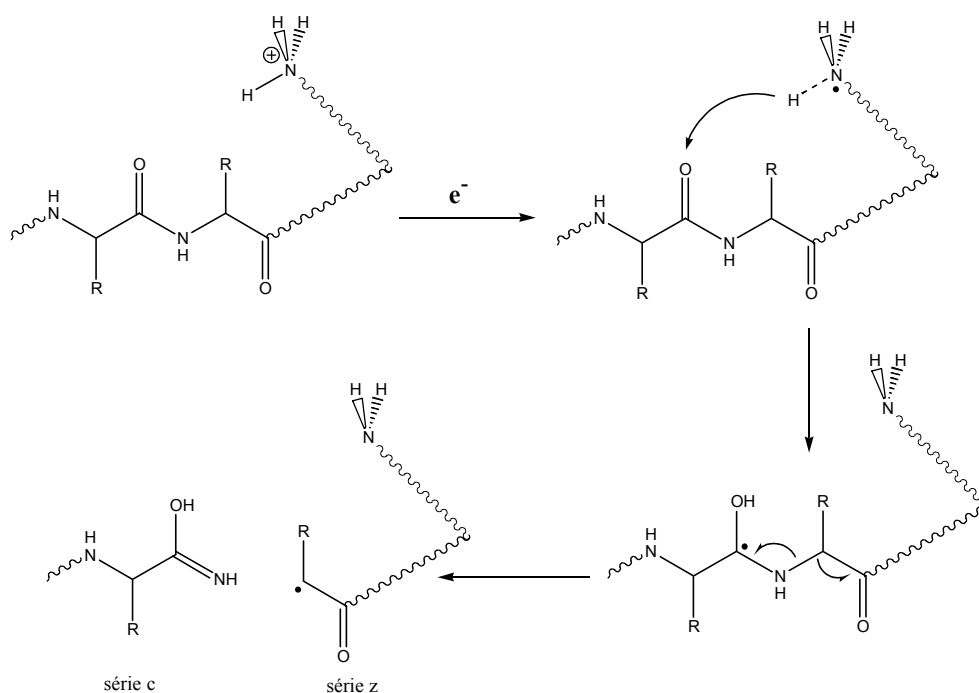
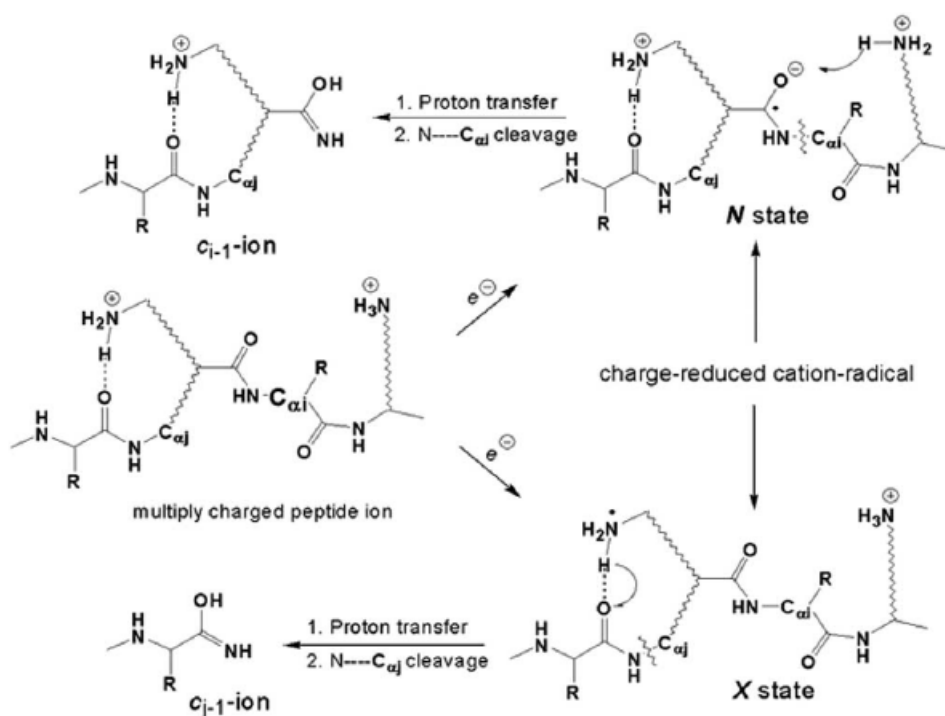


Figure 17: Mécanisme de l'ECD proposé par Zubarev et Mc Lafferty

Puisque cette technique ne provoque pas la rupture des liaisons les plus fragiles du système, la première théorie, proposée par R. Zubarev, est que le mécanisme de capture est non-ergodique, *i.e.* que la rupture de la liaison N-C α a lieu plus rapidement que la redistribution de l'énergie dans tous les niveaux vibrationnels. Cette énergie de recombinaison (ou ER) peut être estimée à environ 4 à 7eV⁵⁷ en fonction de l'état de charge initial de l'ion. Cette valeur est assez faible et correspond pour un ion de 3 kDa à une augmentation de 50-100 K, ce qui est bien inférieur aux 300-600K typiques des techniques d'excitation vibrationnelle. Néanmoins, des calculs théoriques (et des expériences de neutralisation/réionisation) réalisés par F. Tureček^{58,59,60} ne vont pas dans ce sens puisqu'ils montrent que l'attachement d'un radical hydrogène sur un carbonyle conduit à une fragilisation de la liaison N-C α qui peut alors se rompre très facilement pour donner les ions *c* et *z*. Sur la base de ses calculs, F. Tureček propose finalement deux mécanismes : le premier fait intervenir un peptide dans lequel la capture de l'électron se fait par des amides solvatés de manière interne (*Figure 18*).

Figure 18: Mécanismes de l'ECD proposé par Tureček *et al.*

Le second mécanisme repose quant à lui sur la capture de l'électron par un carbonyle d'amide situé loin des charges pour conduire à une espèce de type « super-base » capable de récupérer (lorsqu'elle est dans un état excité approprié) un proton du système et conduire à la rupture de la liaison N-C α . Ce dernier mécanisme permet d'expliquer les nombreuses fragmentations que l'on observe dans les spectres ECD, et notamment celles correspondant à des ruptures de liaisons éloignées des sites de protonation.

D.3.c. Autres techniques de dissociation par FT-ICR

D'autres techniques de dissociation des peptides sont également disponibles sur les instruments FT-ICR. L'une d'entre elle est la dissociation par absorption de plusieurs photons infra-rouge (IRMPD)⁶¹. Les ions piégés dans la cellule ICR sont irradiés par un laser infrarouge (généralement au CO₂). L'absorption hors résonance de plusieurs photons par les ions piégés augmente leur énergie interne jusqu'à conduire à une dissociation. L'un des intérêts majeurs de cette technique est qu'elle ne nécessite pas l'ajout de gaz dans la cellule, permettant de ce fait de maintenir une excellente résolution après la dissociation, sans avoir à attendre plusieurs secondes pour que le gaz ajouté soit pompé. Comme cela a été décrit précédemment, la combinaison ECD/IRMPD est très utile pour l'analyse de protéines

entières⁶². Ceci est rendu possible, dans le cas des instruments Bruker, par l'utilisation d'une cathode émettrice d'électrons annulaire qui permet le passage du laser infrarouge dans la cellule⁶³ (Figure 19).

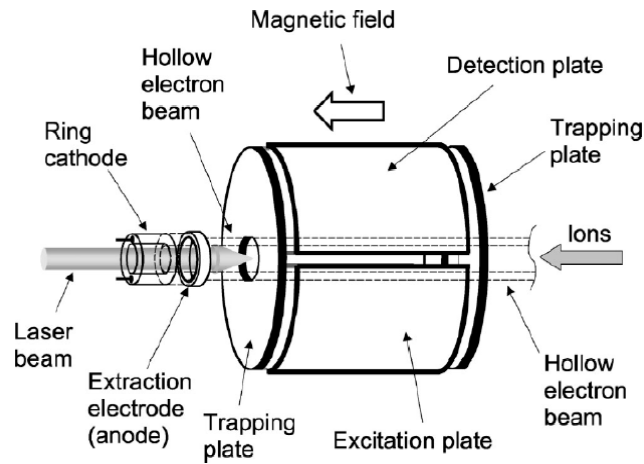


Figure 19: Cellule ICR montrant la possibilité d'utilisation conjointe de l'ECD et de l'IRMPD grâce à un cathode creuse

L'activation des ions peut également s'obtenir par chauffage de la cellule ICR (BIRD)⁶⁴. Les ions piégés vont alors absorber les photons infrarouges émis par rayonnement du corps noir à partir des parois chauffées de la cellule.

Les nouveaux spectromètres hybrides q-FTICR (ou IT-FTICR) ont des capacités au niveau de la fragmentation bien supérieures aux instruments de la génération précédentes (FT-ICR) mais dans ce cas, la fragmentation a lieu dans l'analyseur précédant la cellule ICR qui n'est quasiment plus utilisée que comme cellule de mesure.

Une autre application de l'ECD a été récemment démontrée par l'équipe de R. Zubarev. Ils ont observé que lors d'analyse de peptides par ECD en utilisant des électrons « chauds » (3-13 eV), le mécanisme ECD formait d'abondants ions de type w ⁶⁵, caractéristiques des chaînes latérales des acides aminés. Cette technique appelée HECD pour *Hot Electron Capture Dissociation* permet de distinguer les acides aminés isomères leucine et isoleucine⁶⁶. Cette distinction n'était alors possible par spectrométrie de masse qu'en accédant à de la fragmentation de haute énergie (TOF/TOF, secteurs magnétiques) ou en utilisant différents complexes à base de cuivre (II)⁶⁷ ou de ruthénium (II)⁶⁸.

Un développement récent, lié à l'utilisation de l'ECD est l'approche « top-down » i.e. la fragmentation de protéines entières en phase gazeuse^{69,70}, par opposition aux méthodes classiques d'analyse des protéines appelées « bottom-up ».

- Application au séquençage de protéines par une approche « top-down »

L'approche classique pour identifier une protéine, appelée « bottom-up » (*Figure 20*), repose sur l'analyse, par différentes techniques de spectrométrie de masse, des peptides issus de la digestion enzymatique de la protéine d'intérêt. L'enzyme couramment utilisée est la trypsine qui permet le clivage de la chaîne protéique en position C-terminale des acides aminés de type lysine et arginine.

Une fois la digestion effectuée, deux possibilités sont envisageables. La première est l'établissement d'une cartographie peptidique massique des peptides (souvent par MALDI-TOF). Cette liste de masses, caractéristique de la protéine initiale, est comparée avec les masses des digestions théoriques calculées pour les protéines présentes dans les bases de données via un moteur de recherche comme Mascot (<http://www.matrixscience.com>). Les limitations de cette approche sont multiples : elle est mal adaptée aux mélanges de protéines, nécessite une bonne précision sur la mesure de masse⁷¹ et surtout, le point important, n'est directement applicable qu'aux protéines présentes dans les banques de données.

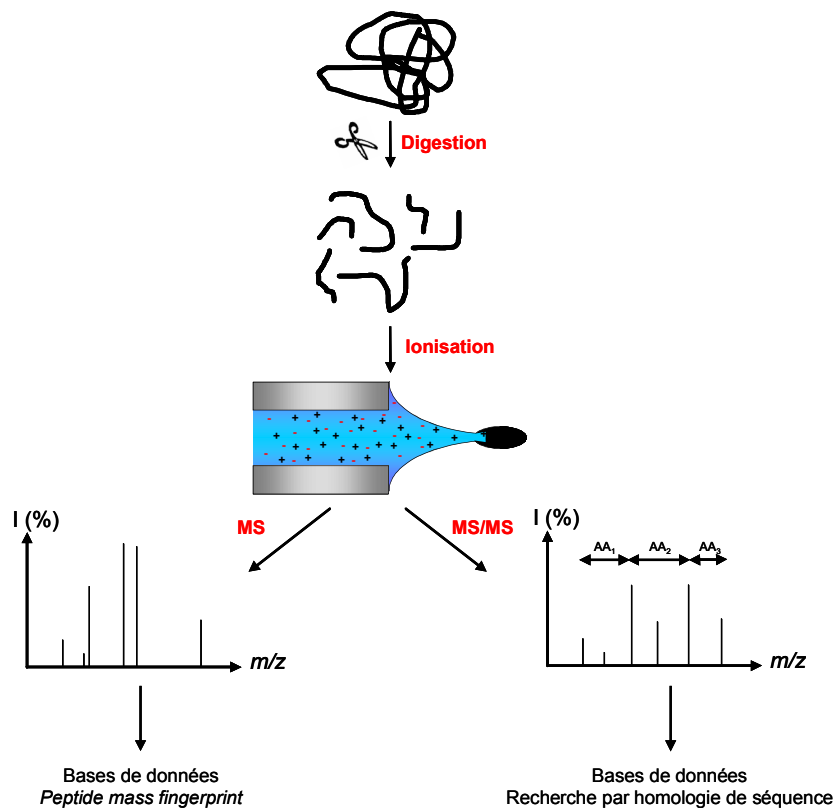


Figure 20: Principe de l'analyse bottom-up en spectrométrie de masse

Après la digestion de la protéine, une seconde approche complémentaire est couramment employée. Elle consiste à séparer les différents fragments de digestion par chromatographie liquide et de fragmenter (MS/MS) le plus de peptides possible. Les spectres de fragmentation bruts obtenus sont comparés aux spectres théoriques des bases de données, ce qui permet de confirmer les résultats obtenus par la cartographie peptidique. De plus, cette approche présente l'avantage de fournir des éléments de séquence importants qui peuvent être exploités par des recherches par homologies de séquences (type Blast).

Plusieurs inconvénients sont liés à l'approche « bottom-up ». Elle nécessite tout d'abord la digestion enzymatique de la protéine, ce qui introduit des étapes de manipulation (et donc de perte de l'échantillon initial). Par ailleurs les digestions à la trypsine nécessitent la présence de résidus basiques (lysine ou arginine) et l'utilisation de plusieurs enzymes est souvent requise pour obtenir une information plus exhaustive sur la séquence de la protéine analysée.

Même si les expériences de MS/MS sur tous les peptides donnent accès à de nombreux *tags*, elles ne permettent ni de trouver les séquences totales de tous les peptides, ni de déterminer leur position exacte au sein de la séquence protéique. Dans ce cas, une (voire plusieurs) nouvelle digestion peut être effectuée sur la protéine avec une enzyme différente qui ne va pas fournir les mêmes peptides. Les *tags* obtenus par fragmentation de ces nouveaux peptides comparés à ceux obtenus lors de la première digestion, permettent dans le meilleur des cas d'améliorer la couverture de séquence obtenue.

La fragmentation des protéines entières en phase gazeuse (ou « top-down ») présente un intérêt croissant pour l'analyse des biomolécules puisqu'elle est en théorie moins consommatrice en échantillon (20-30 pmol) et permet d'accéder rapidement à des informations comme la masse moléculaire de la protéine entière (avec les modifications post-traductionnelles) et des éléments de séquences importants. La fragmentation de composés de haut poids moléculaire implique de pouvoir déterminer l'état de charge élevé des fragments obtenus. De ce fait, l'approche « top-down » nécessite l'utilisation de spectromètres de masse FT-ICR (ou éventuellement certains TOF).

L'approche « top-down » comporte plusieurs étapes (*Figure 21*) :

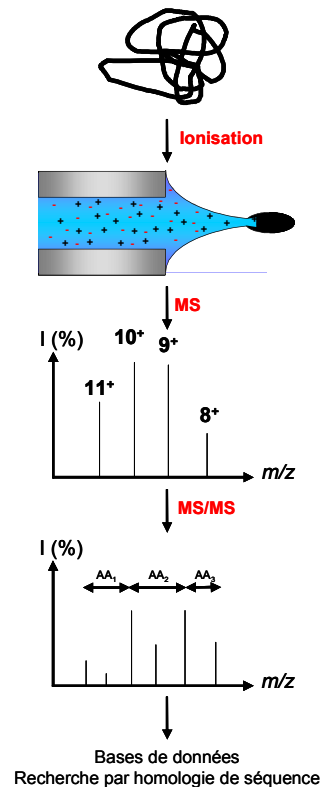


Figure 21: Principe de l'analyse « top-down » par spectrométrie de masse

En premier lieu, il s'agit de mesurer la masse moléculaire de la protéine intacte, c'est-à-dire incluant toutes ses modifications post-traductionnelles. Cette mesure de masse peut être obtenue avec une résolution unitaire. Le record de la masse la plus haute mesurée dans ces conditions est détenu pour la chondroïtinase I dont la masse a été mesurée à 112 509 Da avec une résolution de 170 000⁷². Cette mesure de masse à l'unité près est utile pour identifier la présence de ponts disulfures [S-S vs. 2-SH (2 Da)] ou encore une déamination [-NH₂ vs. -OH (1 Da)], modifications qui n'induisent qu'une très faible variation sur la masse moléculaire.

Ensuite, des informations de séquence peuvent être obtenues par fragmentation de la protéine dans le spectromètre de masse conduisant à son identification dans les banques de données. Différents types d'activation des ions peuvent être utilisés: collision avec un gaz inerte, IRMPD (Infra-Red Multi Photon Dissociation), mais surtout ECD (Electron Capture Dissociation). Ce dernier mode de fragmentation présente un double avantage. Tout d'abord, l'ECD appliqué à une protéine conduit à un très grand nombre de fragments et donc à de nombreuses informations de séquence. Mais surtout, comme nous l'avons vu précédemment, l'ECD laisse intactes toutes les chaînes latérales du squelette polypeptidique. Ainsi, les modifications post-traductionnelles labiles (phosphorylation, O-glycosylation, N-glycosylation,...) restent attachées au squelette ce qui permet leur localisation et leur

identification⁷³. Il faut noter que pour les protéines qui restent structurées en phase gazeuse (en raison principalement du réseau de liaisons hydrogène), c'est généralement la technique d'Activated ECD (AECD) qui est utilisée. Cette approche consiste à « chauffer » la protéine avec des photons IR de faible énergie (par IRMPD) pour casser les liaisons faibles structurant la protéine et pour la déplier en phase gazeuse. Une fois dépliée, l'irradiation par des électrons (ECD) est beaucoup plus efficace.

Les exemples illustrant l'intérêt de combiner différentes techniques de fragmentation pour obtenir d'excellentes couvertures de séquence (>90%) à partir de protéines entières portant des modifications post-traductionnelles variées sont de plus en plus nombreux.^{41,74} N. Kelleher, dont l'équipe est l'une des plus impliquées dans le développement des approches « top-down » a développé un modèle statistique montrant que plus de 99 % des protéines, appartenant à une banque de 5000 protéines différentes, pouvaient être identifiées à partir d'une mesure précise ($\pm 0,1$ Da) de seulement trois ou quatre ions fragments⁷⁵.

Précisons qu'à l'heure actuelle, cette approche ne peut être envisagée que sur les instruments FT-ICR seuls capables d'apporter la résolution et la précision en masse nécessaires à l'interprétation des spectres obtenus pour des molécules de la taille des protéines (résolution des massifs isotopiques complexes obtenus pour les ions fortement chargés).

La mise au point de systèmes « hybrides » *i.e.* combinant la cellule d'analyse FT-ICR avec un analyseur supplémentaire (quadrupôle ou trappe ionique) en amont est un développement nouveau et important pour le développement de stratégies « top-down ». L'implémentation de cet analyseur permet d'une part de simplifier grandement l'utilisation de l'instrument mais aussi d'accroître considérablement ses performances. En effet, il permet d'ajouter une étape de sélection des ions avant leur entrée dans la cellule qui revient à concentrer les ions d'intérêt au détriment des autres. Cette potentialité est d'un intérêt majeur dans le cadre d'une approche top-down sur un mélange de protéines.

On peut néanmoins noter l'utilisation possible, dans certains cas, de spectromètres hybrides Q-TOF pour des analyses « top-down » sur de petites protéines en s'aidant d'outils bioinformatiques pour résoudre les massifs isotopiques des ions fragments plus grand que 9 kDa⁷⁶.

II. Les venins : une source de composés biologiquement actifs

A. Toxines et venins

Les toxines représentent une classe de composés très répandue puisqu'on les trouve chez les animaux et chez les végétaux, mais également chez les bactéries et les microbes. Elles sont utilisées le plus couramment lors d'actes défensifs contre des prédateurs éventuels et aussi lors de processus offensifs comme la capture de proies. Il convient de distinguer les poisons pour lesquels la toxicité résulte d'une ingestion passive des toxines qui sont injectées activement dans l'organisme des proies. Les poisons sont généralement des composés non peptidiques fabriqués par différents organismes dont des algues microscopiques, les dinoflagellés. A titre d'exemple, citons la saxitoxine responsable du syndrome paralytique (*paralytic shellfish poisoning*) qui apparaît chez les consommateurs de coquillages dans lesquelles cette toxine s'accumule. Il a également été montré que certains poissons (comme le Fugu) et certains crabes sont capables d'accumuler la tétródotoxine⁷⁷, fabriquée par une bactérie provenant d'une algue. Certaines grenouilles accumulent quant à elles différents alcaloïdes comme l'épibatidine⁷⁸ à partir de leur régime alimentaire. Plusieurs de ces molécules se sont révélées très intéressantes comme outils pharmacologiques et leur étude a permis de mettre en évidence leur grand potentiel thérapeutique. L'épibatidine et ses dérivés ont ainsi démontré une forte action analgésique due probablement à leur activité sur les récepteurs nicotiniques à l'acétylcholine ce qui a pour conséquence d'intensifier la libération de noradrenaline au niveau de la corde spinale de la moelle épinière^{79,80}.

En revanche, les toxines peptidiques sont généralement trouvées dans les venins animaux et sont liées à un appareil venimeux dont le rôle consiste à injecter le venin dans l'organisme des proies généralement par morsure, piqure ou griffures (*Figure 22*).

Les venins animaux représentent une source très riche de composés biologiquement actifs dont la plus grande partie reste encore à découvrir. Le nombre d'espèces possédant un venin ainsi que leur complexité en terme de contenu peptidique et protéique permet d'évaluer que seulement environ 1% de ces composés bioactifs ont d'ores et déjà été étudiés et caractérisés. Les intérêts majeurs des études menées sur les peptides de venins résident en leur utilisation en recherche biomédicale (fondamentale et diagnostique) et dans le développement de nouveaux médicaments⁸¹.

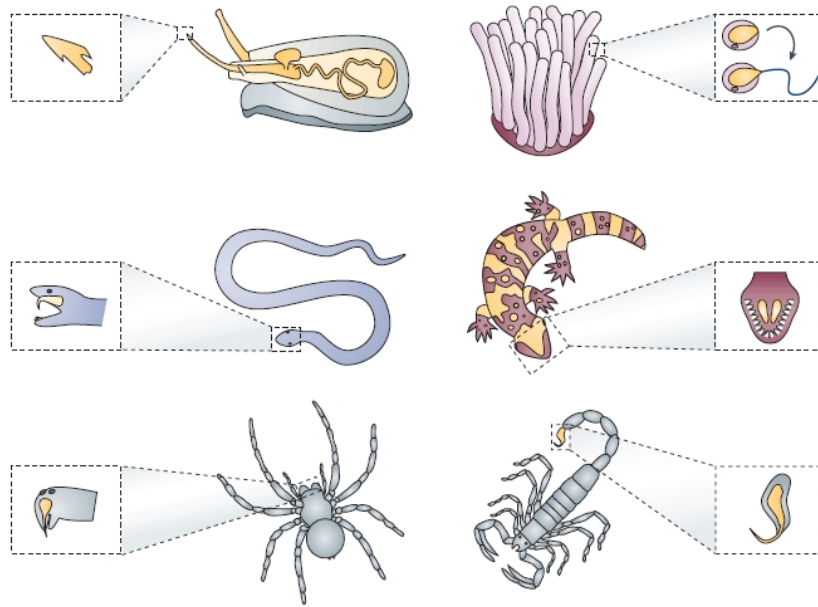


Figure 22: Cônes marins, anémones de mer, serpents, salamandre, araignées et scorpions produisent des peptides aux propriétés thérapeutiques⁸¹

La découverte de nouvelles activités repose en général sur une combinaison de tests biologiques suivis de purifications et d'analyses structurales par différentes méthodes physico-chimiques. Les stratégies d'analyse classiques (dégradation d'Edman des peptides par exemple) et le screening des activités biologiques nécessitent une grande quantité de matériel initial. Elles sont donc plus facilement applicables aux espèces qui peuvent fournir beaucoup de venin comme les serpents mais se révèlent bien plus délicates à mettre en œuvre pour les organismes de petite taille. De plus, afin de prélever le maximum de venin, il est souvent nécessaire de disséquer les glandes à venin ce qui ajoute une nouvelle difficulté lorsque l'on a à faire à des espèces animales relativement rares.

Pour contourner ces problèmes, de nouvelles approches ont été développées. Ainsi, les avancées technologiques comme les séparations par électrophorèse bidimensionnelle ou bien les développements instrumentaux en spectrométrie de masse ont permis de développer considérablement l'étude des peptides et protéines issus de différents milieux biologiques. La toxinologie a ainsi considérablement profité du développement conjoint de la biologie structurale et de la spectrométrie de masse^{82,83}.

De nombreux venins ont déjà fait l'objet d'études extensives : venins d'abeille, de guêpe, de serpent, d'araignée, de scorpion, de cône marin... Dans la plupart des cas, l'objectif est de déterminer de nouvelles activités biologiques potentiellement utilisables en recherche médicale ou pharmacologique.

Dans le cadre de cette thèse, nous avons choisi d'étudier deux types de venins totalement différents : des venins de serpents (*Atractaspis irregularis* et *Dendroaspis angusticeps*) et de cônes marins (*Conus virgo* et *Conus ermineus*). Ainsi, pour des raisons de clarté, seuls les venins de serpents et de cônes marins seront décrits en détail.

A.1. Les venins de serpents

A.1.a. Généralités

Plus de deux mille cinq cents espèces différentes de serpents sont actuellement répertoriées dans le monde. Environ quatre cents d'entre elles sont venimeuses et réputées dangereuses pour l'homme. Ces serpents venimeux sont répartis sur toute la surface du globe. On rencontre les *Elapidae* en Afrique (mambas, cobras), en Asie (cobras, bungares), en Australie (cobras) et en Amérique (serpents corail). Les *Hydrophiidae*, dont la plupart sont totalement adaptés à la vie aquatique, vivent dans les mers tropicales et équatoriales depuis la côte est de l'Afrique jusqu'à la cote ouest de l'Amérique. Ils sont particulièrement abondants dans le sud est asiatique et dans les îlots du Pacifique. Les *Viperidae*, qui abondent en Europe et jusqu'en Asie du sud-ouest, peuvent aussi être rencontrés en Afrique. Les *Crotalidae*, enfin, vivent en Amérique ainsi qu'en Asie.

A l'origine, les venins de serpents proviendraient d'une spécialisation des sécrétions digestives, peut être pancréatiques, certainement salivaires, assurant initialement la digestion des tissus. Le rôle de la salive est double, elle lubrifie les aliments et entame le processus de digestion. Par la suite les serpents auraient développé la capacité de tuer et d'immobiliser leur proie à l'aide de toxines spécialisées pour faciliter leur maintien rendu délicat par l'absence de membres.

A.1.b. Composition des venins de serpent

Les venins de serpents possèdent un contenu protéique important : de l'ordre de 90% du poids sec. Il est classique de séparer les protéines des venins de serpents en deux grandes familles, les enzymes qui ne présentent pas de toxicité aiguë en général, et les toxines dont le rôle pharmacologique est mieux connu.

1- Les enzymes sont généralement des protéines de poids moléculaire plus élevé que les toxines bien que les toxines PLA₂ puissent dépasser les 15 kDa. Certaines enzymes sont

communes à la plupart des venins. C'est le cas notamment des phospholipases qui ont pour rôle d'hydrolyser les phospholipides libres ou membranaires en acides gras et des hyaluronidases qui favorisent la diffusion du venin après son injection. C'est également le cas pour les protéases qui interviennent aussi bien sur les destructions tissulaires observées au cours de nécroses que lors de certains phénomènes pharmaco-toxiques comme les troubles de l'hémostase. S'il est vrai que de nombreuses protéases ne sont pas spécifiques et agissent au niveau de résidus différents, certaines reconnaissent cependant des sites moléculaires particuliers, ce qui en fait des outils très efficaces pour le diagnostic ou le traitement de certaines affections. D'autres enzymes comme l'acétylcholinestérase sont capables d'hydrolyser l'acétylcholine, principal médiateur chimique de l'influx nerveux chez les vertébrés.

2- Les toxines de serpents sont de poids moléculaires variables mais généralement inférieurs à 30 kDa. La plupart d'entre elles ont la propriété de pouvoir se fixer sur un récepteur spécifique, ce qui permet non seulement de se servir de ces toxines comme des sondes pour localiser et quantifier les récepteurs mais aussi de cibler uniquement un type de récepteur bien défini. L'action des toxines peut avoir des conséquences à différents niveaux mais les plus courants sont des effets neurologiques, cardio-vasculaires et musculaires. L'effet toxicologique obtenu va être essentiellement dépendant de la quantité de toxine inoculée et du nombre de récepteurs ciblés ainsi que d'autres facteurs comme la constante d'affinité de la toxine pour son récepteur.

D'un point de vue général, on peut classer les toxines de serpent en huit familles :

- **Les neurotoxines post-synaptiques** peuvent être séparées en trois groupes, les toxines nicotiques, les toxines muscariniques et une troisième famille de toxines regroupant celles qui ne peuvent pas être classées dans les deux premiers groupes. Les toxines nicotiques⁸⁴ (ou curarisantes) agissent sur la membrane post-synaptique et bloquent l'influx nerveux en se fixant sur le récepteur de l'acétylcholine. Elles sont constituées d'une chaîne peptidique de 60 à 74 acides aminés et présentent un poids moléculaire de 7 à 8 kDa. Leur structure tridimensionnelle comporte trois boucles maintenues par la présence de quatre ponts disulfures. On en rencontre trois types, les neurotoxines α -courtes (60-62 acides aminés et 4 ponts disulfures), les neurotoxines α -longues (66-74 acides aminés et 5 ponts disulfures) et les neurotoxines- κ (66-70 acides aminés et 5 ponts disulfures). Les toxines muscariniques⁸⁵ représentent une dizaine de peptides de petite taille (65 acides aminés, 4 ponts disulfures pour une moyenne de 7,5

kDa). Elles présentent une forte affinité avec la partie muscarinique du récepteur cholinergique. La troisième famille de toxines post-synaptiques répertorie les toxines qui ne peuvent pas être classées comme nicotiniques ou muscariniques. Elles sont peu nombreuses comme la vipoxine⁸⁶, isolée du venin de *Daboia russelii* qui n'a aucune action sur les récepteurs nicotiniques et muscariniques mais agit sur les récepteurs adrénergiques, ce qui est exceptionnel chez les serpents.

- **Les neurotoxines présynaptiques**⁸⁷: les toxines présynaptiques bloquantes ou neurotoxines- β sont des entités moléculaires assez complexes puisqu'elles sont composées d'une (comme la notexine), deux (comme la β -bungarotoxine) ou quatre sous-unités (comme la céruléotoxine) mais toutes disposent d'une fonction phospholipase A_2 nécessaire à leur activité toxique. Chaque sous-unité possède de 6 à 8 ponts disulfures pour 13 à 14 kDa de masse moléculaire.

- **Les cardiotoxines** ou cytotoxines regroupent une cinquantaine de peptides dont la structure et la taille sont similaires aux neurotoxines α courtes (60-65 acides aminés, 4 ponts disulfures, conformation en trois boucles). Cependant leur action est totalement différente puisqu'elles dépolarisent rapidement et de manière irréversible la membrane cellulaire, conduisant à sa lyse.

- **Les dendrotoxines**⁸⁸ (ou toxines présynaptiques facilitatrices) ont été découvertes dans les venins des serpents *Dendroaspis* (mambas). Elles font en moyenne 7,5 kDa et présentent trois ponts disulfures. En bloquant les canaux potassium dépendants du potentiel membranaire, elles favorisent la libération d'acétylcholine.

- **Les fasciculines**⁸⁹, également découvertes dans les venins des *Dendroaspis*, présentent 4 ponts disulfures pour environ 7 kDa. Ce sont des inhibiteurs de l'acétylcholinestérase qui s'opposent ainsi à la régulation physiologique de l'influx nerveux.

- **Les myotoxines**⁹⁰ sont des peptides d'environ 5 kDa (3 ponts disulfures) que l'on rencontre dans les venins des *Crotalus*. Elles se fixent sur les canaux ioniques des cellules musculaires et provoquent des nécroses.

- **Les sarafotoxines** représentent une famille de toxines surproduites dans les venins des *Atractaspis*. Elles représentent à elles seules la quasi-totalité de l'activité biologique du venin. Ce sont de petites toxines de 2,5 kDa qui possèdent deux ponts disulfures et qui font partie des espèces les plus toxiques des venins de serpents. Ces puissants peptides vasoconstricteurs sont très proches des endothélines des mammifères sur le plan structural aussi bien que sur le plan fonctionnel.

○ **Les désintégrines**⁹¹ sont trouvées dans les venins de *Viperidae* (5-15 kDa pour 4 à 7 ponts disulfures). Cette famille complexe de toxines inhibe les intégrines, protéines transmembranaires permettant le transfert des messages extracellulaires vers le cytoplasme.

La variabilité des toxines observées dans les venins de serpents a d'ores et déjà permis de découvrir des activités exploitables tant sur le plan pharmaceutique que pharmacologique.

Le Tableau 1 regroupe ces informations.

Toxines	Gamme de masse (kDa)	Nombre de ponts disulfures	Cible biologique et spécificité	Effet
Dendrotoxines	7-8	3	Canaux potassium neuronaux	
Fasciculines	7	4	Inhibiteur de l'acetylcholinestérase	Contractions musculaires brèves, fortes et répétées
Myotoxines	5	3	Inhibiteur des canaux calcium et potassium	Nécrose
Neurotoxine présynaptiques	Plusieurs sous unités de 13 à 14 kDa	6 à 8 par sous-unité	Blocage de la libération de l'acetylcholine - Canaux potassium (β -bungarotoxine)	Action phospholipasique - Paralysie des muscles respiratoires
Sarafotoxines	2-3	2	Récepteurs des endothélines (+ autres?)	Action vasoconstrictrice
Nicotiniques	7-8	4 ou 5	Récepteur cholinergique des fibres musculaires striées	Blocage de l'influx nerveux - Paralysie musculaire
Muscariniques	7-8	4	Partie muscarinique du récepteur cholinergique du système nerveux central	

Tableau 1 : Comparaison (gamme de masse, nombre de ponts S-S, cible biologique, effet) de différentes familles de toxines de venins de serpents

A.1.c. Applications liées à l'utilisation de toxines de serpent

Les exemples d'utilisation de toxines de venins de serpent remontent loin dans le temps puisqu'il semble qu'Aristote incorporait déjà divers extraits de venins, de sang et de viscères de serpents dans les remèdes qu'il fabriquait, à la suite des observations symptomatologiques qu'il avait faites lors d'envenimations.

Les toxines de serpents sont aujourd'hui utilisées dans de nombreux domaines, que ce soit pour la recherche fondamentale ou comme outil de diagnostic ou d'ordre thérapeutique⁹². Sur le plan fondamental, l'utilisation de toxines issues de venin de serpent a permis de grandes avancées dans différents domaines de la biologie :

- Le facteur de croissance des nerfs (NGF), dont l'activité permet de favoriser la multiplication rapide des fibres nerveuses périphériques, a été découvert dans un venin de cobra⁹³. Utilisé pour son activité protectrice des neurones, ce peptide a également été étudié en culture cellulaire, et les études montrent qu'il conduit au développement rapide des synapses et permet d'en expliquer le mécanisme.

- Pour déterminer la structure double hélice de l'ADN, Watson et Crick ont eu recours à des nucléotidases de venin d'un *Naja*. Ces enzymes ont longtemps été utilisées pour déterminer la séquence des bases des différents acides nucléiques avant l'apparition des enzymes de restriction.

- De nombreuses pathologies sont liées au dysfonctionnement des récepteurs à l'acétylcholine (maladie de Parkinson, maladie d'Alzheimer, schizophrénie, névralgies...). Ces pathologies peuvent être explorées grâce à des α -neurotoxines. De plus, les toxines actives sur le canal potassium, comme les dendrotoxines, ont permis de montrer une disparition progressive d'un type précis de canaux potassium (Kv 1.1) au cours du vieillissement biologique. Elles ont également permis de mettre en évidence leur réduction significative et précoce dans la maladie d'Alzheimer, ouvrant la porte à de nouvelles perspectives thérapeutiques.

Les toxines de serpent jouent également un rôle non négligeable dans le développement de nouveaux outils de diagnostic :

- Le ProtacTM est un agent pharmaceutique extrait du venin d'*Agkistrodon contortrix* qui permet d'activer artificiellement une glycoprotéine qui a un double effet anticoagulant, la protéine C⁹⁴. Le test par le ProtacTM permet de vérifier l'intégrité de cette protéine dont les défaillances sont à l'origine d'un grand nombre de thromboses (coagulation *in vivo* d'une artère ou d'une cavité vasculaire ou cardiaque).

- La botrocétineTM, extraite du venin de *Bothrops jararaca*, est un puissant agrégant plaquettaire permettant le diagnostic de plusieurs maladies hémorragiques⁹⁴ d'origine génétique. La maladie de von Willebrandt est caractérisée par le déficit en facteur de von Willebrandt, indispensable à l'adhésion des plaquettes sanguines. La grande qualité de la botrocétine est qu'elle va se fixer à toutes les conformations du facteur de von Willebrandt présentes dans le plasma et non pas aux seules conformations polymériques de haut poids moléculaires comme la ristocétine notamment. Ceci permet donc d'avoir accès à un diagnostic plus précis de la forme de la maladie.

Certaines toxines de serpent sont également utilisées comme de vrais médicaments, aux effets thérapeutiques prouvés :

- La batroxobine (*Bothrops jararaca*) en hydrolysant les liaisons Arg-Gly de la chaîne A alpha du fibrinogène, entraîne la libération du fibrinopeptide A, mais n'entraîne pas la libération du fibrinopeptide B. Les conséquences observées sont une diminution du taux de

fibrinogène dans le sang et de la viscosité sanguine, réduisant de ce fait les risques de thromboses (formation de caillots de sang).

- Les toxines de serpent semblent également très prometteuses pour les traitements contre le cancer. Par exemple, l'éristostatine⁹⁵, extraite de *Eristocophis macmahoni*, permet d'induire une destruction des cellules des métastases du mélanome (apoptose) sans action cytotoxique directe. La contortrostatine⁹⁶, extraite du venin du crocodile *Agkistrodon contortrix*, permet d'inhiber l'adhésion des cellules cancéreuses aux protéines de structure, ce qui permet de les empêcher de traverser les membranes basales et de bloquer leur dispersion dans l'organisme.

Enfin, un exemple frappant de médicament créé à partir d'une toxine du venin est le développement du Captopril[®]. La toxine Teprotide a initialement été extraite du venin du serpent *Bothrops jararaca*⁹⁷ et s'est révélée efficace pour le traitement de l'hypertension artérielle et les insuffisances cardiaques. La voie empruntée par les inventeurs du Captopril[®] pour le développement du médicament est donnée en Figure 23.

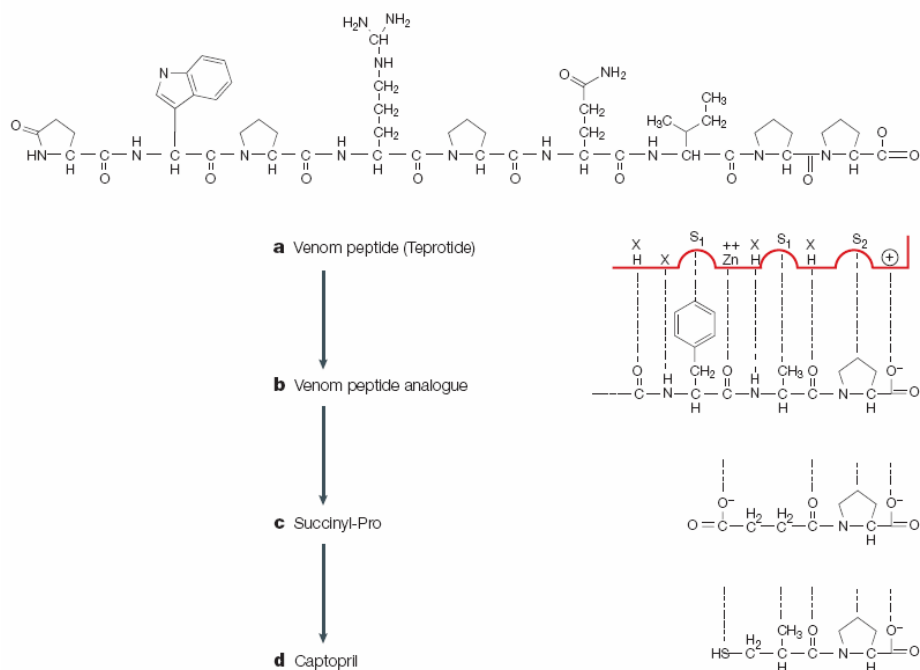


Figure 23: Le développement du Captopril[®] à partir d'une toxine du venin du serpent *Bothrops jararaca*⁸¹

(a) Teprotide, octapeptide extrait du venin de *Bothrops jararaca*, inhibiteur de l'enzyme de conversion de l'angiotensine (ACE).

(b) Relation structure-activité (SAR) du teprotide avec le site de liaison de l'ACE.

(c) Premier composé synthétisé à partir des données pharmacologiques obtenues par les études de SAR. Le problème de ce composé est qu'il développe une faible affinité avec le site de liaison malgré une très bonne sélectivité pour l'ACE.

(d) Le Captopril, médicament ingérable par voie orale, le groupement thiol qui remplace le carboxylate de (c) permet d'obtenir une meilleure affinité tout en maintenant la bonne sélectivité.

Enfin, il ne faut pas oublier l'utilisation des venins de serpent en immunologie. Les chevaux auxquels des doses contrôlées de venin de serpent sont injectées produisent des anticorps leur permettant de lutter contre la toxicité du venin injecté. Pour obtenir un antivenin (antidote), il suffit d'effectuer des prises de sang chez le cheval immunisé, d'en extraire les anticorps qui peuvent alors être injectés chez une personne mordue par un même serpent, ce qui est le traitement le plus efficace en cas de morsure grave. Vital Brazil, l'un des fondateurs du célèbre Institut Butantan au Brésil, fut le premier en 1900 à immuniser des chevaux dans le but de produire des antivenins monovalents (un seul type de venin est utilisé pour immuniser les chevaux) et polyvalents (un mélange de venins de différents serpents est utilisé lors de l'immunisation devant servir à traiter les cas de morsures dues à des serpents brésiliens. L'Institut Butantan est aujourd'hui encore l'un des plus grands centres de production de sérums antivenimeux (serpents, scorpions, araignées, etc.).

A.2. Les venins de cônes marins

A.2.a. Cônes marins et mode d'envenimation

Décrit pour la première fois en 1758 par Linné, le genre *Conus* appartient à la famille des *Conidae*, dans la classe des mollusques gastéropodes. Il existe au moins 700 espèces de cônes, principalement répartis dans les eaux chaudes de la région Indo-Pacifique, de l'Atlantique Sud et des Caraïbes. De nouvelles espèces sont découvertes chaque année. Les cônes se caractérisent par une grande variété de tailles, de formes, de couleurs et de motifs arborés par leur coquille (*Figure 24*) :



Figure 24: Quelques coquilles de cônes marins

Ce sont tous des prédateurs carnivores qui peuvent être classés en trois catégories suivant leur régime alimentaire. Il existe ainsi les cônes vermivores qui se nourrissent de vers, des cônes malacophages qui se nourrissent de mollusques ou encore des cônes piscivores. Il est difficilement compréhensible que ces mollusques, de morphologie générale voisine de l'escargot, puissent être de redoutables prédateurs! En fait, ils disposent d'un mode d'envenimation aussi ingénieux qu'efficace. Tout commence au niveau de leur appareil digestif qui est caractérisé par un ruban de dents (*radula*) plus ou moins évolué. Les dents, d'une longueur variant de 0,1 mm à 10 mm sont en forme de flèche creuse. Leur morphologie est étroitement liée au régime alimentaire des cônes. Par exemple, la dent des espèces piscivores est un véritable harpon (Figure 25) permettant au cône de contrôler le poisson lors du choc excitotoxique provoqué par la piqure.

L'appareil venimeux des cônes marins est assez complexe (Figure 25). Il possède tout d'abord une glande musculaire (*vb pour venom bulb*) qui ne secrète pas le venin mais dont le rôle serait uniquement mécanique. Elle permettrait par contraction, d'éjecter le venin vers le proboscis (*pr*). Cette glande se prolonge par le conduit à venin (*vd pour venom duct*) dans lequel le venin est secrété et stocké. Ce dernier peut être très long puisque à titre d'exemple un cône de type *Conus textile* de 9 cm peut présenter un conduit d'une longueur de 25 cm⁹⁸. Le conduit à venin débouche au niveau du pharynx (*p*). Chaque cône possède un sac radulaire (*rs pour radula sac*) composé de deux parties distinctes. La première partie, longue, permet de stocker les dents en cours de développement tandis que la seconde partie, plus courte, abrite les dents radulaires prêtes à l'emploi. Il est important de noter que chacune des dents fabriquées n'est utilisée qu'une seule et unique fois. La réserve est de ce fait primordiale pour la survie du

cône. Chacune de ces dents se termine par un véritable harpon, qui permet non seulement de créer une plaie importante mais également de maintenir la proie accrochée à la dent. Le sac radulaire se prolonge par le pharynx, liaison entre le conduit à venin, le *proboscis* et le tube digestif. Le *proboscis* est l'organe clé du mécanisme d'envenimation puisqu'il est non seulement capable de fixer un harpon à son extrémité directement à partir du sac radulaire mais également de se déplier vers la cible afin de faire pénétrer la dent remplie de venin dans la chair de la proie.

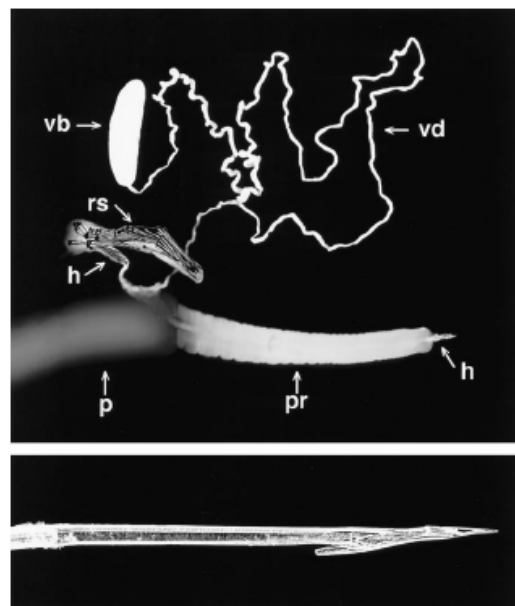


Figure 25: Appareil venimeux et dent de type harpon de *conus purpurascens*⁹⁹

Le mode d'envenimation des cônes est très sophistiqué et l'on peut séparer les cônes en deux grandes classes. Une première classe utilise le *proboscis* pour se munir d'un harpon issu du sac radulaire et pour le projeter vers la proie, qui peut être située à plusieurs centimètres. D'un coup bref, le cône enfonce la dent remplie de venin dans les tissus de la victime. La proie est alors paralysée et gobée. Un deuxième mode opératoire utilise les capacités de leur bouche très extensible pour emprisonner la proie. L'envenimation s'effectue dans un second temps par le même principe du harpon. Ces deux techniques d'envenimations sont très bien illustrées dans la Figure 26 par B. Olivera⁹⁹ :

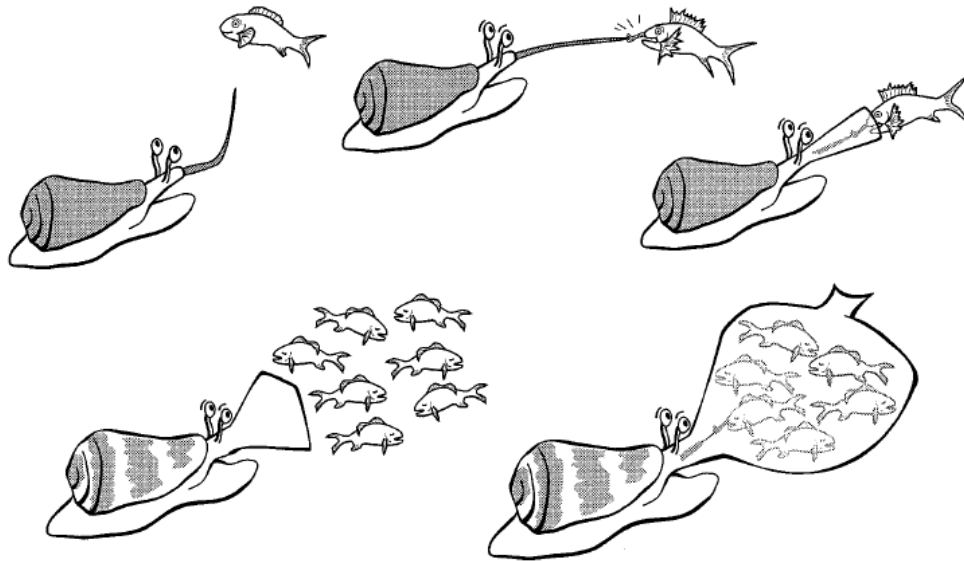


Figure 26: techniques d'envenimation des cônes marins⁹⁹

Les cas d'envenimation chez l'homme, bien que rares, peuvent être mortels. Les effets de la piqûre, souvent douloureuse, se caractérisent par un engourdissement et un gonflement rapide et localisé autour de la plaie. La paralysie s'étend peu à peu à tout le corps, accompagnée selon les cas de diplopie et de nausées. Dans les cas les plus graves, le décès survient par arrêt cardiaque ou respiratoire.

A.2.b. Conopeptides et conotoxines

Bien que la toxicité des cônes ait été mentionnée depuis plus de 300 ans, le venin des cônes ne suscite la curiosité des scientifiques que depuis le début des années 1970. Comme la plupart des animaux venimeux, les cônes utilisent leur venin pour capturer des proies, pour se protéger et se défendre. Ce sont de redoutables prédateurs dont le venin est très complexe et susceptible de contenir plusieurs centaines de molécules bio actives dont de nombreux petits peptides très structurés : les conopeptides. On estime à 50.000 le nombre de conopeptides présents dans les venins des différents cônes. Ces peptides ont une longueur de chaîne variable allant de 6 à 41 acides aminés, mais la majorité possède entre 12 et 30 acides aminés¹⁰⁰. L'une des caractéristiques structurales des conopeptides est que, malgré une taille assez réduite, ils possèdent une forte structuration conformationnelle. La présence d'un nombre important de ponts disulfures dans leur séquence est la principale explication à cette structuration mais il est aussi montré que certains conopeptides présentent des hélices α qui permettent de stabiliser des modifications post-traductionnelles, très nombreuses au sein de cette famille^{101,102,103}.

Le terme « conotoxine » peut être appliqué à tous les conopeptides réticulés par au moins deux ponts disulfures. Ces molécules présentent donc une structure très compacte et sont de ce fait très stables et très résistantes aux attaques enzymatiques pouvant être déclenchées après une envenimation. Il existe une nomenclature pour les conotoxines basée sur l'activité biologique, le nom du cône et le motif cystéine retrouvé dans la séquence. L'ensemble des conotoxines peut également être divisé en « superfamilles » puis en « familles ». Les peptides d'une même « superfamille » partagent un même agencement des cystéines et une séquence signal du précurseur très conservée alors que les peptides d'une même « famille » ont en commun leur activité pharmacologique. A ce jour, on dénombre 7 superfamilles de conotoxines: A, M, O, P, S, T, I auxquelles viennent s'ajouter conopressines, conantokines, contryphans, conorfamides et contulakines.

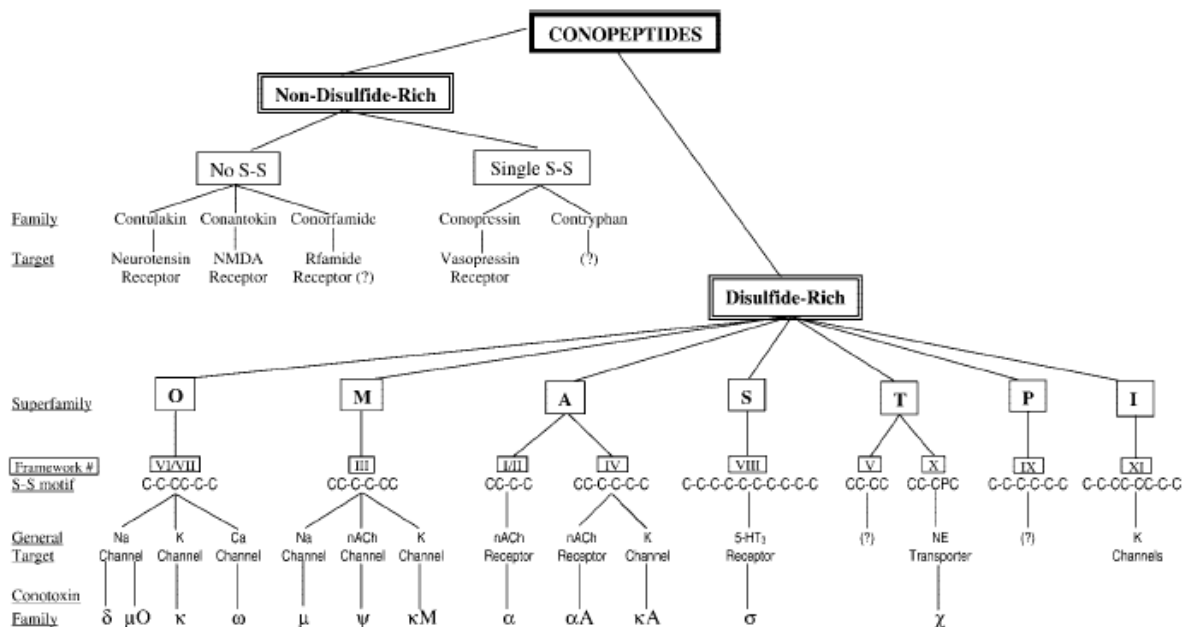


Figure 27: Diagramme d'organisation des peptides de cônes indiquant la superfamille d'appartenance, l'organisation des ponts disulfures et les cibles pharmacologiques connues¹⁰⁴

D'un point de vue pharmacologique, chaque peptide lie spécifiquement une sous-unité d'un récepteur ou d'un canal ionique. Il peut s'agir des récepteurs couplés à la protéine G, des canaux ioniques dépendant du potentiel ou des canaux ioniques dépendant d'un ligand. Ces propriétés font des conotoxines des outils pharmacologiques extrêmement intéressants pour l'étude des récepteurs et particulièrement des récepteurs orphelins.

A.2.c. Développement de nouveaux ligands et de nouveaux médicaments à partir des venins de cônes marins

- *Les conotoxines comme ligands des canaux ioniques*

Les canaux ioniques dépendant du potentiel sont des assemblages de protéines membranaires possédant la particularité de changer de conformation en fonction de la valeur du potentiel transmembranaire. Lors de la variation de ce dernier, ils vont s'ouvrir pour laisser passer des ions ou bien se fermer pour empêcher tout passage constituant ainsi de véritables pores pour les cellules. Ces protéines ont une sélectivité différente pour les cations ou les anions et sont de ce fait couramment divisées en quatre grands groupes : les canaux sodium (Na^+), potassium (K^+), calcium (Ca^{2+}) et chlore (Cl^-). Ces canaux ioniques ont pour principal rôle physiologique la génération, la formation et la transduction des signaux électriques des cellules. Ils sont la cible de toxines de différents organismes.

Toxines liant les canaux ioniques dépendant du voltage

Les canaux sodium dépendant du potentiel sont des molécules clés pour la génération et la propagation de potentiels d'action de cellules électriquement excitables. Ils causent la dépolarisation de la membrane en permettant à un flux d'ions sodium de pénétrer dans la cellule. Il est aujourd'hui évalué qu'environ 7000 ions sodium passent à travers le canal à chaque fois que ce dernier est activé (période de l'ordre d'une milliseconde). Ce canal n'a jamais pu être complètement caractérisé par cristallographie du fait de sa taille et de sa nature hydrophobe (*Figure 28*). Il est formé de quatre domaines transmembranaires et chacun d'entre eux est divisé en plusieurs sous-unités α . Aujourd'hui, dix isoformes de sous-unités α ont été identifiées et classifiées sous une nomenclature unique de type $\text{Na}_v1.x$, où x représente la sous-unité en question (donc variant de 1 à 10). Conventionnellement, ces canaux sont divisés en deux classes pharmacologiques, l'une sensible à la tetrodotoxine (TTX) et l'autre non-sensible à la TTX, la TTX étant la toxine classiquement utilisée pour bloquer les canaux Na_v et dont le site d'interaction avec le pore est appelé site I.

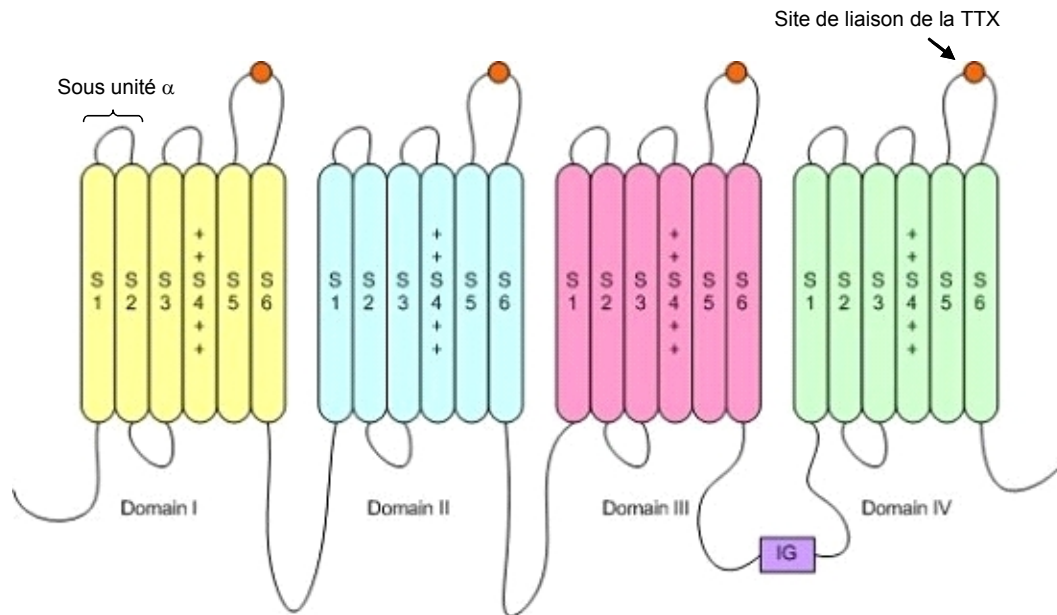


Figure 28: Schéma d'un canal ionique sodique dépendant du potentiel

Trois familles de conotoxines sont connues pour lier les canaux sodiques dépendant du potentiel. Il s'agit des μ -conotoxines qui vont bloquer totalement le canal, des μ O-conotoxines qui vont inhiber sa conductance et des δ -conotoxines qui vont en retarder voire en inhiber l'inactivation rapide. Les μ -conotoxines (Figure 27) bloquent les courants Na^+ en agissant directement sur le site I, site d'interaction de la TTX. Jusqu'à présent, ces conotoxines sont les seuls peptides connus à agir sur ce site d'interaction. Il semblerait que pour bloquer le pore, elles agissent non seulement comme une barrière stérique mais aussi comme une barrière électrostatique¹⁰⁵. Les μ O-conotoxines se rapprochent des μ -conotoxines car elles inhibent également la conductance des canaux Na^+ , mais le mécanisme d'action est différent. En effet, pour les μ O-conotoxines, il n'y a pas de compétition observée avec la saxitoxine (STX) pour la liaison avec le canal. Bien que le mécanisme d'interaction ne soit pas encore bien défini, ceci démontre le fait que ces toxines n'interagissent pas avec le site I du canal¹⁰⁶. Les μ O-conotoxines et les δ -conotoxines se ressemblent de par leur grande hydrophobicité qui doit jouer un rôle important dans les mécanismes de liaison aux canaux. Les effets des δ -conotoxines dépendent réellement du système étudié. Par exemple, la δ -TxVIA (aussi appelée peptide King Kong), extraite du *Conus textile* prolonge la durée du courant de sodium chez les mollusques alors que chez les vertébrés, elle se fixe sur les canaux sodiques sans aucun effet toxique.

Les canaux potassium sont responsables de la repolarisation de la cellule après un potentiel d'action mais ont également une action variée sur différents types de cellules¹⁰⁷. Cette grande diversité d'action est illustrée par le fait que pas moins de 80 gènes encodent différents canaux potassiques qui peuvent être divisés en deux sous-classes. La première est constituée des canaux qui possèdent six domaines transmembranaires (K_v) et la seconde de ceux qui n'en possèdent que deux (K_{ir}).

Les conotoxines liant les canaux K^+ commencent seulement à être étudiées. La première, isolée du *conus purpurascens* est la κ -conotoxine PVIIA¹⁰⁸. Le blocage du canal par cette toxine est physiologiquement intéressant et constitue une action clé pour une immobilisation rapide de la proie. Il est intéressant de noter qu'elle possède également une action synergique avec la δ -conotoxine PVIA conduisant à une hyperexcitation de l'animal envenimé. Le résultat se traduit par une paralysie (tétanie) instantanée. D'autres conotoxines sont connues pour se lier spécifiquement aux canaux potassiques dépendant du potentiel, il s'agit notamment des κA - et κM -conotoxines. Les κA -conotoxines sont des peptides O-glycosylés et plusieurs résultats sur l'activité de la κA -SIVA (*Conus striatus*) ont montré qu'elle serait impliquée dans le blocage du canal K^+ . Cependant, des études complémentaires sur cette famille et sur des peptides similaires devront être menées afin d'en tirer de réelles conclusions¹⁰⁴. Les κM -conotoxines bloquent également les canaux K^+ avec une affinité de l'ordre de 1 μM pour la κM -RIIIK¹⁰⁹ (*Conus radiatus*).

Les canaux calcium sont uniques parmi les différents canaux ioniques par leur implication dans la dépolarisation de la membrane, et leur couplage avec le déclenchement de processus physiologiques tels que la libération de neurotransmetteurs et l'activation des gènes. Ils sont classés en fonction de leurs propriétés électrophysiologiques et pharmacologiques en six groupes (types L, N, P, Q, T et R). A titre d'exemple, les canaux de type N ($Ca_v2.2$) sont surexprimés dans la moelle épinière en conditions de douleur chronique et sont impliqués dans les voies nociceptives afférentes. L'une des approches qui a permis de mettre en évidence leur implication dans ces processus est l'injection locale ou systémique d'inhibiteurs pharmacologiques spécifiques. Les plus spécifiques d'entre eux sont les ω -conotoxines GVIA¹¹⁰ et MVIIA¹¹¹ extraites respectivement des venins de *Conus geographus* et *Conus magus*. L' ω -conotoxine GVIA est probablement la conotoxine la plus utilisée en neuroscience puisque plusieurs centaines de publications rapportent son utilisation comme outil pharmacologique pour l'inhibition de la transmission synaptique.

Toxines liant les canaux dépendant d'un ligand

Les canaux ioniques dépendant d'un ligand sont des complexes protéiques transmembranaires permettant ou non la circulation d'ions au niveau de la membrane cellulaire. Contrairement aux canaux dépendant du potentiel, ils ne sont pas activés par une différence de potentiel mais par une molécule organique. Trois grands groupes de canaux peuvent être caractérisés. Le premier, le plus important, regroupe les canaux dérivant de la même famille de gènes, et sont activés par l'acétylcholine, la sérotonine, l'acide gamma amino-butyrique (GABA) ou la glycine. Le second groupe comporte des canaux qui dérivent d'une seconde famille et que l'on rassemble sous le nom de récepteurs au glutamate. Ils sont généralement divisés en deux sous-familles qui séparent ceux qui sont sensibles au N-méthyl-D-aspartate (NMDA) et ceux qui y sont insensibles. Le troisième et dernier groupe de canaux ioniques dépendant d'un ligand et impliqués dans la transmission synaptique regroupe les récepteurs à l'ATP. Structuralement tous ces canaux ioniques sont constitués de cinq domaines qui possèdent chacun quatre hélices α transmembranaires.

Au niveau de la toxinologie, des conopeptides se liant aux trois familles de canaux ioniques ligand-dépendants ont été identifiés. De nombreuses conotoxines, issues de différents venins, lient les récepteurs nicotiniques à l'acétylcholine¹¹². La plupart d'entre elles vont entrer en compétition avec les ligands endogènes, ce sont les agonistes. Cependant, des antagonistes non compétitifs ont également été répertoriés. Une famille particulière de conopeptides, les conantokines, a révélé des affinités particulières pour les sous-types de récepteurs au glutamate sensibles au NMDA. Cette famille se caractérise notamment par la présence d'acide γ -carboxyglutamate et une absence de pont disulfures. La famille des α -conotoxines est la plus largement représentée au sein des venins de cônes. Ce sont des antagonistes très sélectifs de sous-types du récepteur nicotinique à l'acétylcholine. Une autre toxine attire l'intérêt en neuroscience. Il s'agit de la σ -conotoxine GVIIIA¹¹³ qui lie spécifiquement le récepteur 5-HT₃, sensible à la sérotonine. Il s'agit d'un peptide de 41 acides aminés et possédant 10 cystéines reliées par 5 ponts disulfures. Parmi les acides aminés présents dans sa séquence se trouve un tryptophane bromé qui semble simuler le ligand endogène, la sérotonine étant un dérivé hydroxylé du tryptophane.

- *Les conotoxines, médicaments du futur ?*

Dans le développement de nouveaux médicaments à partir des toxines, il semblerait que les cônes marins aient un rôle à jouer important (*Tableau 2*).

Conopeptide	Conus Species (*)	Target	Stage	Company **	Other name(s)	Comment	Reference
α -Conotoxin Vc1.1	<i>Conus victoriae</i> (m)	Competitive blocker of selected neuronal-type nicotinic ACh receptors	Preclinical	Metabolic Pharmaceuticals, Melbourne, Australia	ACV1	Effective against peripheral neuropathic pain in animal models and accelerates functional recovery of injured neurons	[16, 29]
rho- Conotoxin T1A	<i>Conus tulipa</i> (p)	Acts as reversible noncompetitive inhibitor of α -1 adrenergic receptors	Preclinical	Xenome, Ltd., Brisbane, Qld., Australia		Acts as a reversible noncompetitive inhibitor of alpha-1 adrenergic receptors	[30]
ω -conotoxin CVID	<i>Conus catus</i> (p)	Blocks N-type calcium channel specific sub-type.	Stage II	AMRAD Operations under licence from Univ. of Queensland	AM336	Reported to have a better therapeutic index than Prialt TM	[31]
ω -conotoxin MVIIA	<i>C. magus</i> (p)	N-type calcium channels	Stage III	Elan Corporation (Elan Pharmaceuticals), CA, USA	SNX-III, C1002, Ziconotide, Prialt TM	Significant pain relief to patients in clinical trials. Side effects in some patients. Hence call for repeat of Stage III clinical trials for cancer pain.	[32-35]
χ -conopeptides (χ -CTX MrIA/B)	<i>Conus marmoreus</i> (m)	Acts as reversible noncompetitive inhibitor of the neuronal noradrenaline transporter	Preclinical	Xenome, Ltd., Brisbane, Qld., Australia		Being developed to "treat certain types of pain, for which there is currently a lack of effective treatment" neuropathic pain.	[30]
Contulakin-G	<i>Conus geographus</i> (p)	Binds to neurotensin receptor	Stage II	Cognetix Inc, Salt Lake City, USA	CGX-1160	Short term management of post-operative pain	[36]
Conantokin-G	<i>Conus geographus</i> (p)	Selective inhibitor of the NMDA receptor (NR2B subtype)	Stage II	Cognetix Inc, Salt Lake City, USA	CGX-1007	Potent antinociceptive effects in several models of injury-induced pain. Also, control of seizures in intractable epilepsy	[36]
Conantokin-T	<i>Conus tulipa</i> (p)	Selective inhibitor of the NMDA receptor (NR2A and NR2B) subtypes	Stage II	Cognetix Inc, Salt Lake City, USA	CGX-100	Potent antinociceptive effects in several models of injury-induced pain	[(36]

Tableau 2: Présentation de différentes conotoxines en cours de développement clinique¹¹⁴

En effet, l'utilisation de certaines α -conotoxines ou ω -conotoxines s'est révélée prometteuse dans le traitement de la douleur ou de certaines maladies neurodégénératives. Les canaux calcium de type N régulent la libération de neurotransmetteurs au niveau des synapses centrales et périphériques. Une forte densité de ces canaux est présente dans la moelle épinière où arrivent les terminaisons des fibres nociceptives A δ et C. Cette observation a permis d'envisager l'utilisation d'antagonistes des canaux sodium de type N tels que l' ω -conotoxine MVIIA pour supprimer la douleur. En plus de son efficacité supérieure à celle des opioïdes, l' ω -conotoxine MVIIA ne présente aucun effet de désensibilisation ni augmentation de tolérance contrairement à la morphine. Après autorisation de la FDA (Food and Drug Administration), ce peptide a été commercialisé aux Etats-Unis sous le nom générique de ziconotide et sous le nom commercial de Prialt[®], par la société Elan Pharmaceuticals. Une autre ω -conotoxine, CVID de *C. catus* est en développement clinique pour ses propriétés antinociceptives.

Les α -conotoxines ciblant les récepteurs nicotiniques neuronaux peuvent également être utiles dans le traitement de maladies liées à des dérégulations de ces récepteurs : maladie d'Alzheimer, maladie de Parkinson, schizophrénie, douleur...

Le peu d'espèces caractérisées au regard de la richesse des venins indique que le potentiel à exploiter est grand et que de nombreuses molécules aux activités biologiques variées restent encore à découvrir.

B. Stratégies classiques d'analyse des venins

Chaque analyse de venin nécessite une stratégie adaptée. De par les différences qui existent d'un venin à l'autre, des quantités de matériel disponible et de l'objectif que l'on souhaite atteindre (caractérisation totale du venin, d'une famille de toxines, d'une toxine unique, d'une activité biologique etc...), aucune stratégie standardisée n'est applicable à l'ensemble des cas de figure. D'une manière générale, l'analyse d'un venin nécessite une combinaison de plusieurs techniques biochimiques et physico-chimiques complémentaires, voire même parfois de biologie moléculaire. Dans ce paragraphe, sont détaillées les stratégies préférentiellement utilisées pour la caractérisation ciblée de toxines et pour la recherche de ligands spécifiques de récepteurs. Dans chacun de ces cas, l'apport de la spectrométrie de masse sera explicité.

B.1. Cartographie d'un venin brut

Etablir la cartographie (massique, chromatographique ou électrophorétique) d'un venin permet de créer une empreinte unique caractéristique de ses différents constituants. Plusieurs techniques permettent d'atteindre cet objectif. Les techniques électrophorétiques sur gel de polyacrylamide (mono ou bidimensionnel) en font partie.

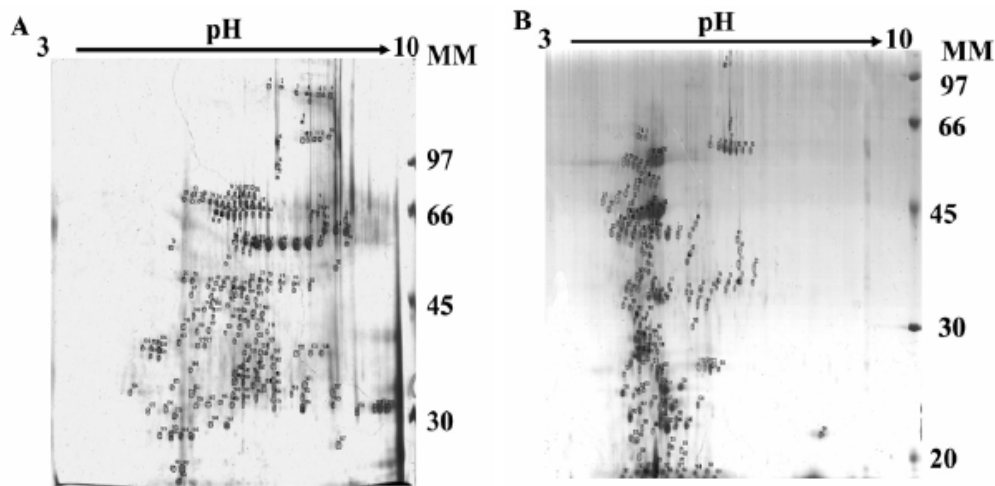


Figure 29: Cartographie électrophorétique de deux venins de serpents par approche 2D-PAGE – (A) *Naja naja atra* et (B) *Agkistrodon halys*¹¹⁵

La Figure 29 montre la séparation par électrophorèse bidimensionnelle de deux venins de serpents. Il apparaît clairement que les empreintes obtenues sont totalement différentes. Le *Naja naja atra* présente des toxines sur une large gamme de pH (4 à 9 unités) pour des masses dépassant 100 kDa tandis que les toxines de l'*Agkistrodon halys* sont centrées sur une gamme de pH plus étroite (entre 4 et 7) pour des masses ne dépassant pas 66 kDa.

On peut également réaliser des cartographies massiques (plus résolutive) par spectrométrie de masse que ce soit par ESI-QTOF¹¹⁶, MALDI-TOF-MS¹¹⁷ ou par l'utilisation d'un couplage entre la chromatographie liquide (polarité de phase inversée) et la spectrométrie de masse (MS ou MS/MS)¹¹⁸.

L'approche par MALDI-TOF semble être la technique de choix pour l'analyse des venins bruts¹¹⁹. Elle permet d'obtenir principalement des ions monochargés ce qui rend les analyses beaucoup plus simples à exploiter. Il est connu que cette méthode induit une certaine discrimination des familles de composés au niveau de l'étape de désorption-ionisation. Pour limiter ce phénomène de suppression ionique, le choix de la matrice est important et, afin d'obtenir le maximum d'informations, différents spectres acquis avec différentes matrices sont souvent nécessaires. L'analyse de venins bruts par ESI-MS n'a été effectuée que plus récemment¹²⁰ et ce mode d'ionisation est principalement employé en couplage avec la chromatographie liquide. Une limitation de l'analyse directe est souvent la présence de sels dans les venins. Ceci entraîne un double problème : d'une part une forte concentration en sels va induire une forte instabilité du spray²⁵ et la présence d'adduits va abaisser le signal du peptide d'intérêt (déjà divisé par le nombre d'états de charges observé).

L'intérêt de cartographier un venin ne se situe pas qu'au niveau moléculaire mais peut également être utilisé en taxonomie. En effet, une large proportion des composés présents dans les venins est constituée de peptides et de protéines, molécules provenant de l'expression directe du génome. Ainsi, ces cartographies peuvent s'apparenter à une « image » partielle du génome des espèces étudiées, permettant de rapprocher des spécimens sur le plan génétique. L'utilisation des profils des venins à des fins taxonomique a été démontrée pour les bactéries¹²¹, les venins de serpents¹²², de cônes¹¹⁸, de scorpions¹²³ et d'araignées^{124,125}. Toutes ces applications montrent que l'analyse des venins par spectrométrie de masse représente une technique rapide et fiable d'identification et de détermination de sous-classes d'espèces morphologiquement voisines.

Récemment, P. Escoubas *et al.*¹²⁶ ont développé le concept de cartographies de venins en trois dimensions. Dans la première dimension, on trouve la masse moléculaire, dans la seconde, le temps de rétention (caractère hydrophobe) et dans la dernière, l'intensité des pics dans les spectres de masse MALDI-TOF. Au final, le graphique obtenu donne une bonne représentation des toxines constituant un venin et permet de cibler spécifiquement une toxine d'intérêt en fonction de sa localisation dans le paysage en 3D.

B.2. Recherche de toxines d'intérêts

Parmi les dizaines, voire les centaines de constituants d'un venin, un défi est d'aller directement rechercher des toxines d'intérêt ciblé, c'est-à-dire dont l'activité de la famille d'appartenance est déjà connue. Un exemple est celui de la recherche de sarafotoxines (peptides vasoconstricteurs) dans le venin d'*Atractaspis irregularis*, dont des isoformes avaient déjà été caractérisées dans trois autres venins. Cette étude fait l'objet d'un chapitre de cette thèse.

La difficulté réside dans le fait de cibler les molécules d'intérêt dans le mélange complexe. Alors que le génome humain est entièrement séquencé, aucun génome d'animal venimeux ne l'est, ce qui complique fortement l'analyse protéomique des venins. Néanmoins, il semblerait que le futur proche de l'analyse des venins animaux passe par ce type d'étude appelée « vénomique »¹²⁷. Ce projet ambitieux consiste d'un côté en l'analyse du génome de spécimens sélectionnés et de l'autre en l'analyse des peptides et protéines présents dans le venin ainsi que des ARN messagers présents dans les cellules des glandes à venins. Ainsi, même si l'analyse du génome de ces espèces n'est pas encore disponible, le clonage moléculaire peut être employé pour rechercher des familles de toxines dans les venins.

B.2.a. Recherche d'une activité biologique par des tests de liaison

Il est possible de rechercher une toxine pour (et par) son activité biologique. La stratégie la plus couramment employée pour étudier l'activité des différents constituants d'un venin comporte de nombreuses étapes. Comme nous l'avons précédemment vu, le venin est très complexe et une première étape de fractionnement par chromatographie est nécessaire. Chacune des fractions obtenues est alors testée sur divers modèles biologiques afin de déterminer si elle contient ou non une activité intéressante (*Figure 30*). A titre d'exemple, une activité spécifique et sélective de récepteurs pour lesquels aucun ligand n'est connu serait particulièrement attractive.

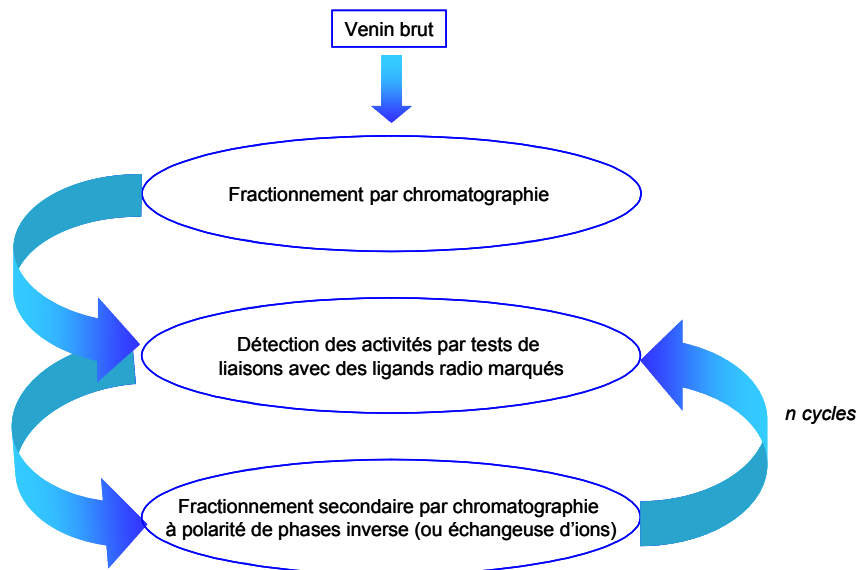


Figure 30: Stratégie de découverte de nouvelles activités pharmacologiques employés pour le venin de *Dendroaspis angusticeps*

L'activité biologique est visualisée par des tests de liaison qui emploient des ligands radioactifs marqués. La fraction est mise en présence d'un récepteur et d'un de ses ligands connus marqué par radioactivité. Si le ligand froid (composant du venin) entre en compétition avec le ligand radioactif pour sa liaison avec le récepteur, cela modifie le signal radioactif mesuré en fin d'expérience. Cette technique permet de démontrer non seulement qu'un ligand froid se lie à un récepteur mais aussi d'en étudier l'affinité. Ainsi en testant chacune des fractions sur différents modèles ligand radioactif/récepteur, il est possible de déterminer si la fraction contient une toxine qui se lie sur les récepteurs testés.

La mesure de radioactivité de fin d'expérience est comparée à deux références : le BT (*Binding total*) et le NS (*Non Specific*). La première correspond à la mesure à l'équilibre de la radioactivité du mélange récepteur et ligand radioactif, sans ajout de fraction. Ceci nous permet d'accéder à la valeur maximale de radioactivité que le système puisse fournir puisque le maximum de ligand radioactif va se lier à ses récepteurs cibles sans qu'une compétition avec un autre composé ne soit possible. La seconde référence utilisée correspond à la mesure de la radioactivité du système après avoir mis une forte quantité de fraction au contact des récepteurs et du ligand marqué. Comme nous nous trouvons très loin de l'équilibre, seule une très faible quantité de ligand radioactif est liée aux récepteurs. La radioactivité mesurée correspond donc uniquement à la liaison non spécifique (NS) du ligand radioactif aux récepteurs (*Figure 31*).

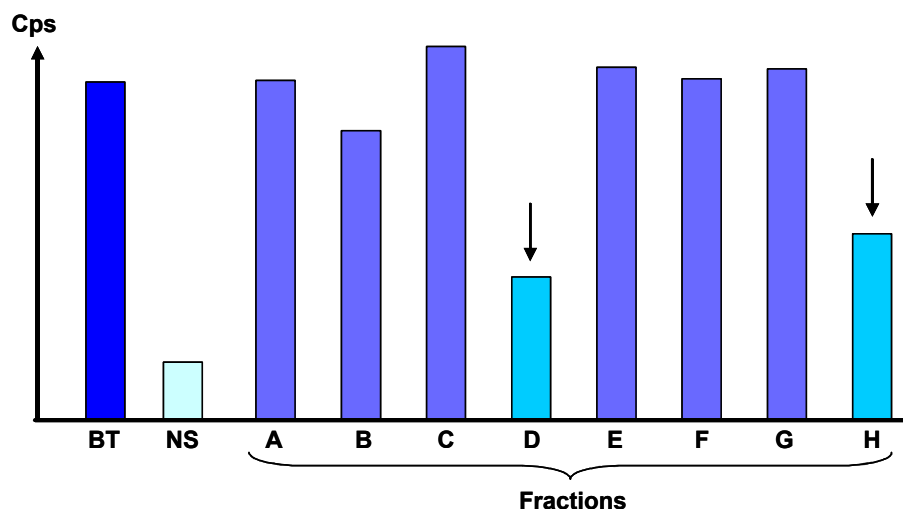


Figure 31: Exemple de présentation des résultats d'expériences de liaisons effectuées sur 8 fractions pour un modèle récepteur/ligand radioactif donné- BT= Liaison totale, NS= Liaison non spécifique – le signal mesuré est donné en cps : « nombre de coups par seconde » relatif au comptage radioactif

Sur la *Figure 31*, il apparaît clairement que seules deux fractions contiennent des molécules qui entrent en compétition avec le ligand radioactif et qui entraînent une baisse de signal (fractions D et H). Toutes les autres fractions présentent un signal de l'ordre de celui du *Binding Total*, ce qui signifie qu'un maximum de ligand radioactif s'est fixé au récepteur et qu'aucun composé n'est venu perturber cette liaison.

Dans le cas où l'un des tests est jugé satisfaisant (comme pour D et H dans notre exemple), la fraction sélectionnée est fractionnée de nouveau mais par un autre processus chromatographique afin de mieux séparer les différents constituants de cette dernière. Chacune des fractions secondaires est testée de nouveau afin de retrouver l'activité biologique

et ainsi de déterminer quelle fraction la contient. L'étape de purification/vérification de l'activité est répétée jusqu'à l'obtention d'un unique pic chromatographique actif.

L'électrophysiologie est également très utilisée en différents modes (patch-clamp, voltage clamp) pour non seulement déterminer les cibles moléculaires des toxines mais aussi afin de comprendre leur mode d'action. Ces informations sont accessibles par l'analyse de divers paramètres électrophysiologiques, comme des mesures de potentiels de membranes ou de réponses synaptiques.

La suite de la stratégie est basée sur la caractérisation structurale de la toxine active. De nombreuses techniques peuvent être utilisées. La dégradation d'Edman (voir en Annexe), la résonance magnétique nucléaire et la spectrométrie de masse sont les techniques les plus employées. En général, la structure de toxines est déterminée par une combinaison de ces trois techniques^{128,129,130} puisque chacune de ces méthodes est complémentaire. La dégradation d'Edman permet d'avoir accès à la structure primaire de toxines isolées. Cependant, son efficacité dépend de la pureté de la toxine, un unique pic chromatographique n'est pas toujours caractéristique d'une molécule unique. De plus, cette approche est très délicate lorsque peu de matériel est disponible et inefficace dans le cas où l'extrémité *N*-terminale est modifiée. La résonance magnétique nucléaire est la technique la plus résolutive des trois citées puisqu'elle permet d'avoir accès à la structure tridimensionnelle de toxines séquencées. Son défaut majeur est qu'elle nécessite une quantité d'échantillon importante bien que cette contrainte soit de moins en moins marquée. Enfin, la spectrométrie de masse est la plus sensible des trois techniques et permet d'accéder aux séquences primaires des toxines *via* des digestions protéolytiques pour les toxines de haut poids moléculaire.

B.2.b. Apport du clonage moléculaire

Le clonage moléculaire est une technique devenue routinière en biologie moléculaire qui consiste à cloner des fragments d'ADN dans différents vecteurs plasmidiques. Lors de la transcription chaque type de cellule va, à partir de ses gènes, synthétiser les ARN messagers (ARNm) nécessaires à son fonctionnement. Ainsi, chacune des cellules disposera au final d'un ensemble d'ARN messagers différents, spécifiques de sa fonction. Par exemple, une partie des ARN messagers présents dans les cellules des glandes à venins des serpents va être responsable de la biosynthèse de l'ensemble des molécules protéiques et peptidiques retrouvées dans le venin, dont de nombreuses toxines. L'étude de ces ARN est donc

primordiale en particulier pour comprendre le mécanisme de maturation des toxines dans les glandes à venin. Les différentes étapes qui permettent de passer de l'ADN génomique à l'obtention des séquences peptidiques caractérisant des toxines sont détaillées en annexe.

Cependant, tous les ARNm clonés à partir des glandes à venin ne correspondent pas à des séquences de toxines. Deux difficultés apparaissent donc lorsque l'on dispose des seuls résultats de clonage. La première est de savoir si les séquences nucléotidiques obtenues correspondent bien à des séquences de toxines effectivement présentes dans le venin. Pour répondre à cette question, il est possible, dans certains cas, d'utiliser des amorces spécifiques (pour la RT-PCR) des toxines recherchées, ce qui permet de n'amplifier que le matériel génétique codant pour ces toxines. La deuxième difficulté vient du fait qu'il n'est pas aisé de déterminer la séquence exacte des toxines déduites des ARNm. En effet, même si le début de la traduction est déterminée par la méthionine d'initiation et la fin par le codon stop, la séquence reliant ces deux marqueurs n'est pas forcément celle de la toxine que l'on retrouve dans le venin. L'intervention d'endopeptidases qui vont modifier la longueur de ces toxines jusqu'à l'obtention de la forme mature et leurs modifications post-traductionnelles ne peuvent être déterminées qu'en analysant réellement le venin au niveau moléculaire.

Dans ce contexte, la spectrométrie de masse a un rôle important à jouer. La comparaison des masses calculées pour les séquences (de longueur variable) obtenues par clonage moléculaire à celles obtenues lors de la cartographie massique du venin permet de relier rapidement les données biologiques aux analyses par spectrométrie de masse. Le fait que la FT-ICR permette des mesures de masses très précises est un avantage considérable dans cette recherche. Néanmoins, des expériences de spectrométrie de masse en tandem à partir d'autres instruments sur les peptides sélectionnés peuvent confirmer toute ou partie de la séquence de ces derniers. Ces expériences sont dans la plus part des cas réalisées (en ligne) après séparation des différents constituants par chromatographie liquide.

B.2.c. Rôle de la spectrométrie de masse

Le rôle de la spectrométrie de masse pour caractériser précisément une toxine d'intérêt s'est affirmé ces dernières années. On se retrouve, pour l'analyse de ces toxines, dans des cas où les génomes ne sont pas séquencés ; il s'agit donc d'un vrai séquençage *de novo*. Pour les peptides de haut poids moléculaire ou les petites protéines, l'approche classique par digestion enzymatique suivie de l'analyse des fragments peptidiques en MS/MS est généralement ce qui

est utilisé. Pour les petits peptides purifiés, diverses expériences de fragmentation peuvent être envisagées (basse énergie, haute énergie) pour obtenir le plus d'informations possibles concernant les séquences. Dans la plupart des cas, des instruments de types tQ ou Q-TOF sont utilisés car ils permettent d'obtenir des spectres MS/MS de qualité. Assez peu d'études font état de l'utilisation de la spectrométrie de masse FT-ICR pour analyser des toxines animales¹³¹. C'est dans ce domaine que se situe notre travail puisque nous avons, au cours de cette thèse, développé plusieurs stratégies d'analyse de venins bruts basées sur l'utilisation de la spectrométrie de masse FT-ICR et notamment appliqué la stratégie « top-down » à une toxine d'intérêt.

Références bibliographiques

- ¹ Thomson J.J., Cathodes rays, **1897**, *Philosophical Magazine*, 44, 293-316.
- ² Thomson J.J., On the masses of the ions in gases at low pressures, **1899**, *Philosophical Magazine*, 48, 547-567.
- ³ Aston F.W., The constitution of the elements, **1919**, *Nature*, 104, 393.
- ⁴ Aston F. W., Isotopes and atomics weights, **1920**, *Nature*, 105, 617.
- ⁵ Dempster A.J., A new method of positive ray analysis, 1918, *Physical Reviews*, 11, 316-324.
- ⁶ Wiley W.C. and McLaren I.H., Time of Flight mass spectrometer with improved resolution, **1955**, *The review of scientific instruments*, 26 (12), 1150-1157.
- ⁷ Paul W., Steinwedel H., A new mass spectrometer without magnetic field, **1953**, *Z. Naturforsch.*, 8A, 448-450.
- ⁸ Comisarow M.B. and Marshall A.G., Fourier transform ion cyclotron resonance mass spectroscopy, **1974**, *Chemical Physics Letters*, 25, 282-283.
- ⁹ Makarov A., Electrostatic axially harmonic orbital trapping: A high-performance technique of mass analysis, **2000**, *Anal. Chem.*, 72, 1156-1162.
- ¹⁰ Munson M.S.B., Field F.H., Chemical Ionization Mass Spectrometry I. General introduction, **1966**, *J. Am. Chem. Soc.*, 88 (12), 2621-2630.
- ¹¹ Benninghoven A., Loebach E., Tandem Mass Spectrometer for Secondary Ion Studies, **1971**, *Review of Scientific Instruments*, 42 (1), 49-52.
- ¹² Macfarlane R.D., Torgerson D.F., Californium-252 plasma desorption mass spectroscopy, **1976**, *Science*, 191(4230), 920-925.
- ¹³ Barber M., Bordoli R.S., Sedgwick R.D., Tyler, A.N., Fast atom bombardment of solids as an ion source in mass spectroscopy, **1981**, *Nature*, 293, 270-275.
- ¹⁴ Karas M., Bachmann D., Bahr U., Hillenkamp F., Matrix-Assisted ultraviolet laser desorption of non-volatile compounds, **1987**, *J. Mass Spectrom. Ion Processes*, 78, 53-68.
- ¹⁵ Karas M., Hillenkamp F., Laser desorption ionization of proteins with molecular mass exceeding 10,000 Daltons, **1988**, *Analytical Chemistry*, 60, 2299-2301.
- ¹⁶ Tanaka K., Waki H., Ido Y., Akita S., Yoshida Y., Yoshida T. Protein and polymer analysis up to m/z 100,000 by laser ionization time-of-flight mass spectrometry , **1988**, *Rapid Commun. Mass Spectrom.*, 2, 151-153.
- ¹⁷ Fenn J.B., Mann M., Meng C.K., Wong S.F., Whitehouse C.M., Electrospray Ionization for Mass Spectrometry of Large Biomolecules, **1989**, *Science*, 246, 64-71.
- ¹⁸ Dole R.L. Mack L.L., Hines R.L., Mobley R.C., Ferguson L.D., Alice M.B., Molecular beams of macroions, **1968**, *Journal of Chemical Physics*, 49 (5), 2240-2249.
- ¹⁹ Zeleny J, The electrical discharge from liquid points, and a hydrostatic method of measuring the electric intensity at their surfaces, **1914**, *Physical Reviews*, 3, 69-91.
- ²⁰ Kebarle P., A brief overview of the present status of the mechanisms involved in electrospray mass spectrometry, **2000**, *J. Mass Spectrom.*, 35, 804-817.
- ²¹ Taylor G.I., Disintegration of water drops in an electric field, **1969**, *Proceedings of the Royal Society in London A*, 383-397.
- ²² Rayleigh Lord, On the equilibrium of liquid conducting masses charged with electricity, *Philosophical Magazine*, **1882**, 14, 184-186.

- ²³ Gomez A, Tang K., Charge and fission of droplets in electrostatics sprays, **1994**, *Physics of Fluids*, 6, 404-414.
- ²⁴ Iribarne J.V., Thomson B.A., On the evaporation of small ions from charged droplets, **1976**, *Journal of Chemical Physics*, 64, 2287-2294.
- ²⁵ Cole R.B., Some tenets pertaining to electrospray ionization mass spectrometry, **2000**, *J. Mass Spectrom.*, 35, 763-772.
- ²⁶ Kebarle P, Peschke M, On the mechanisms by which the charged droplets produced by electrospray lead to gas phase ions, **2000**, *Anal. Chim. Acta*, 406, 11-35.
- ²⁷ Fenn J. B., Ion formation from charged droplets: Role of geometry, energy and time, **1993**, *J. Am. Soc. Mass Spectrom.*, 4, 524-535.
- ²⁸ Fernandes de la Mora J., Electrospray ionization of large multiply charged species proceeds via Dole's charged residue mechanism, **2000**, *Anal. Chim. Acta*, 406, 93-104.
- ²⁹ Fenn J.B., Rosell J., Nohmi T., Banks Jr F.J., **1995**, In *Biochemical and Biotechnological Applications of Electrospray Ionisation Mass Spectrometry*, Snyder A.P. (ed), American Chemical Society, Washington DC, 650-680.
- ³⁰ Wilm M., Mann M., Electrospray and Taylor-Cone theory, Dole's beam of macromolecules at last?, **1994**, *Int. J. Mass Spectrom. Ion Processes*, 136, 167-180.
- ³¹ Wilm M., Mann M, Analytical properties of the nanoelectrospray ion source, **1996**, *Anal. Chem.*, 68, 1-8.
- ³² Jurashek R., Dülcks T. and Karas M., Nanoelectrospray – More than just a minimized-flow electrospray ionisation source, **1999**, *J. Am. Soc. Mass Spectrom.*, 10, 300-308.
- ³³ Li Y, Cole R. B., Shifts in peptide and protein charge state distributions with varying spray tip orifice diameter in nanoelectrospray Fourier transform ion cyclotron resonance mass spectrometry, **2003**, *Anal. Chem.*, 75, 5739-5746.
- ³⁴ Patrie S.M., Charlesbois J.P., Whipple D., Kelleher N.L., Hendrickson C.L., Quinn J.P., Marshall A.G., Mukhopadhyay B., Construction of a hybrid quadrupole/Fourier transform cyclotron resonance mass spectrometer for versatile MS/MS above 10 kDa, **2004**, *J. Am. Soc. Mass Spectrom.*, 15, 1099-1108.
- ³⁵ Petermann S.M., Dufresne C.P., Horning S., The use of a hybrid linear trap/FT-ICR mass spectrometer for on line high resolution/ high mass accuracy bottom-up sequencing, **2005**, *J. Biol. Tech.*, 16, 112-124.
- ³⁶ Guan S., Marshall A.G., Sheppele S.E. Resolution and chemical formula identification of aromatic hydrocarbons containing sulfur, nitrogen, and/or oxygen in crude oil distillates, **1996**, *Anal. Chem.*, 68, 46-71.
- ³⁷ Hughey C.A., Rodgers R.P., Marshall A.G. Resolution of 11000 compositionally distinct components in a single electrospray ionization Fourier Transform Ion Cyclotron Resonance mass spectrum of crude oil, **2002**, *Anal. Chem.*, 74, 4145-4149.
- ³⁸ Cooper H.J., Marshall A.G., Electrospray ionization Fourier transform mass spectrometric analysis of wine, **2001**, *J. Agric. Food Chem.*, 49, 5710-5718.
- ³⁹ Smith R.D., Anderson G.A., Lipton M.S. Pasa-Tolic L., Shen Y., Conrads T.P., Veenstra T.D., Udseth H.R., An accurate mass tag strategy for quantitative and high-throughput proteome measurements. **2002**, *Proteomics*, 2, 513-523.
- ⁴⁰ Jones J.J., Stump M.J., Fleming R.C., Lay J.O.Jr, Wilkins C.L., Investigation of MALDI-TOF and FT-MS techniques for analysis of *Escherichia coli* whole cells, **2003**, *Anal. Chem.*, 75, 1340-1347.

- ⁴¹ Ge Y., Lawhorn B.G., ElNaggar M., Strauss E., Park J-H., Begley T.P., McLafferty F.W., Top-down characterization of larger proteins (45 kDa) by electron capture dissociation mass spectrometry, **2002**, *J. Am. Chem. Soc.*, 124, 672-678..
- ⁴² Sleno L., Volmer D.A., Ion activation methods for tandem mass spectrometry, **2004**, *J. Mass Spectrom.*, 39, 1091-1112.
- ⁴³ Laskin J., Denisov E.V., Shukla A.K., Barlow S.E., Futrell J.H., Surface-induced dissociation in a Fourier transform ion cyclotron resonance mass spectrometer: instrument design and evaluation, **2002**, *Anal. Chem.*, 74, 3255-3261.
- ⁴⁴ Biemann K., Martin S.A., Mass spectrometric determination of the amino acid sequence of peptides and proteins, **1987**, *Mass Spectrom. Rev.*, 6, 1-76.
- ⁴⁵ Dongre A.R., Jones J.L., Somogyi A., Wysocki V.H., Influence of peptide composition, gas-phase basicity and chemical modification on fragmentation efficiency: Evidence of the mobile proton model, **1996**, *J. Am. Chem. Soc.*, 118, 8365-8374.
- ⁴⁶ Wysocki V.H., Tsaprailis G., Smith L.L., Brei L.A., Mobile and localized protons: a framework for understanding peptide dissociation, **2000**, *J. Mass Spectrom.*, 35, 1399-1406.
- ⁴⁷ Roepstorff P., Fohlman J., Proposal for a common nomenclature for sequence ions in mass-spectra of peptides, **1984**, *J. Biomed. Mass Spectrom.*, 11, 601.
- ⁴⁸ Paizs B., Suhai S., Fragmentation pathways of protonated peptides, **2005**, *Mass Spectrom. Rev.*, 24, 508-548.
- ⁴⁹ Biemann K., Four decades of structure determination by mass spectrometry : from alkaloids to heparin, **2002**, *J. Am. Soc. Mass Spectrom.*, 13, 1254-1272.
- ⁵⁰ Senko M.W., Speir J.P., McLafferty F.W., Collisional activation of large multiply charged ions using Fourier Transform Mass spectrometry, 1994, *Anal. Chem.*, 66, 2801-2808.
- ⁵¹ Marshall A.G., Hendrickson C.L., Stone D., Shi H., Scaling MS plateaus with high resolution FT-ICRMS, **2002**, *Anal. Chem.*, 75, 253A-259A.
- ⁵² Bossio R.E., Marshall A.G., Baseline resolution of isobaric phosphorylated and sulfated peptides and nucleosides by electrospray ionisation FTICR ms: another step toward mass-spectrometry based proteomics, **2002**, *Anal. Chem.*, 74, 1674-1679.
- ⁵³ Zubarev R.A., Kelleher N.L., McLafferty F.W., Electron capture dissociation of multiply charged protein cations. A Nonergodic Process., **1998**, *J. Am. Chem. Soc.*, 120, 3265-3266.
- ⁵⁴ Stenballe A., Jensen O.N., Olsen J.V., Haselmann K.F., Zubarev R.A., Electron capture dissociation of singly and multiply phosphorylated peptides, **2000**, *Rapid Commun. Mass. Spectrom.*, 14, 1793-1800.
- ⁵⁵ Mirgorodskaya E., Roepstorff P., Zubarev R.A., Localization of O-glycosylation sites in peptides by electron capture dissociation in a Fourier transform mass spectrometer, **1999**, *Anal. Chem.*, 71, 4431-4436.
- ⁵⁶ Zubarev R.A., Kruger N.A., Fridriksson E.K., Lewis M.A., Horn D.M., Carpenter B.K., McLafferty F.W., Electron capture dissociation of gaseous multiply-charged proteins is favoured at disulfide bonds and other sites of high hydrogen atom affinity, **1999**, *J. Am. Chem. Soc.*, 121, 2857-2862.
- ⁵⁷ Zubarev R.A., Haselmann K.F., Budnik B., Kjeldsen F., Jensen F., Towards an understanding of the mechanism of electron capture dissociation: a historical perspective and modern ideas, **2002**, *Eur. J. Mass Spectrom.*, 8, 337-349.

- ⁵⁸ Turecek, F., N-C α bond dissociation energies and kinetics in amide and peptide radicals. Is the dissociation a non-ergodic process?, **2003**, *J. Am. Chem. Soc.* 125, 5954-5963.
- ⁵⁹ Turecek, F and Syrstad, E.A. Mechanism and energetics of intramolecular hydrogen transfer in amide and peptide radicals and cation-radicals, **2003**, *J. Am. Chem. Soc.*, 125, 3353-3369
- ⁶⁰ Syrstad E.A., Turecek F., Toward a general mechanism of electron capture dissociation, **2005**, *J. Am. Soc. Mass. Spectrom.*, 16, 208-224.
- ⁶¹ Little D.P., Speir J.P., Senko M.W., O'Connor P.B., McLafferty F.W., Infrared multiphoton dissociation of large multiply charged ions for biomolecules sequencing, **1994**, *Anal. Chem.*, 66, 2809-2815.
- ⁶² Hakansson K., Chalmers M.J., Quinn J.P., McFarland M.A., Hendrickson C.L., Mashall A.G., Combined electron capture and infrared multiphoton dissociation for multistage MS/MS in a Fourier transform ion cyclotron mass spectrometer, **2003**, *Anal. Chem.*, 75, 3256-3262.
- ⁶³ Tsyboin Y.O., Witt M., Baykut G., Kjeldsen F., Hakansson P., Combined infrared multiphoton dissociation and electron capture dissociation with a hollow electron beam in Fourier transform ion cyclotron resonance mass spectrometry, **2003**, *Rapid Commun. Mass Spectrom.*, 17, 1759-1768.
- ⁶⁴ Price W.D., Schnier P.D., Williams E.R., Tandem mass spectrometry of large biomolecule ions by blackbody infrared radiative dissociation, **1996**, *Anal. Chem.*, 68, 859-866.
- ⁶⁵ Kjeldsen F., Haselmann K. F., Budnik B.A., Jensen F., Zubarev R.A., Dissociative capture of hot (3-13eV) electrons by polypeptide polycations: an efficient process accompanied by secondary fragmentation, **2002**, *Chem. Phys. Lett.*, 356, 201-206.
- ⁶⁶ Kjeldsen F., Haselmann K.F., Sorensen E.S., Zubarev R.A., Distinguishing of Ile/Leu amino acid residues in the PP3 protein by (hot) electron capture dissociation in Fourier transform ion cyclotron resonance mass spectrometry, **2003**, *Anal. Chem.*, 75, 1267-1274.
- ⁶⁷ Wee S., O'Hair R. A. J., McFadyen D.W., Side-chain radical losses from radical cations allows distinction of leucine and isoleucine residues in the isomeric peptides Gly-XXX-Arg, **2002**, *Rapid Commun. Mass Spectrom.*, 16, 884-890.
- ⁶⁸ Ito A., Okamura T., Yamamoto H., Ueyama N., Ake K., Masui R., Kuramitsu S., Tsunasawa S., Distinction of Leu and Ileu using a ruthenium(II) complex by MALDI-LIFT-TOF/TOF-MS analysis, **2005**, *Anal. Chem.*, 77, 6618-6624.
- ⁶⁹ Kelleher N.L., Top-Down proteomics, **2004**, *Anal. Chem.*, 76, 197A-203A.
- ⁷⁰ Reid G.E., McLuckey S.A., "Top-down" protein characterization via tandem mass spectrometry, **2002**, *J. Mass Spectrom.*, 37, 663-675.
- ⁷¹ Clauser K.R., Baker P., Burlingame A.L., Role of accurate mass measurement (± 10 ppm) in protein identification strategies employing MS or MS/MS and database searching, **1999**, *Anal. Chem.*, 71, 2871-2882.
- ⁷² Kelleher N.L., Senko M.W., Siegel M.M., McLafferty F.W., Unit resolution mass spectra of 112 kDa molecules with 3 Da accuracy, **1997**, *J. Am. Soc. Mass. Spectrom.*, 8, 380-383.
- ⁷³ Kelleher N.L., Costello C.A., Begley T.P., McLafferty F.W., Thiaminase I (42 kDa) Heterogeneity, Sequence Refinement, and Active Site Location from High-Resolution Tandem Mass Spectrometry, **1995**, *J. Am. Soc. Mass Spectrom.*, 6, 981-984.

- ⁷⁴ Ge Y., ElNaggar M., Sze S.K., Oh H.B., Begley T.P., McLafferty F., Top Down characterization of secreted proteins from mycobacterium tuberculosis by electron capture dissociation mass spectrometry, **2003**, *J. Am. Soc. Mass Spectrom.*, 14, 253-261.
- ⁷⁵ F. Meng, B. J. Cargile, L.M. Miller, J. R. Johnson, and N. L. Kelleher, Informatics and multiplexing of intact protein identification in bacteria and the archaea, *Nat. Biotechnol.*, **2001**, 19, 952-956
- ⁷⁶ Nemeth-Cawley J.F., Tangarone B.S., Rouse J.C., “Top Down” characterization is a complementary technique to peptide sequencing for identifying protein species in complex mixtures, **2003**, *J. Proteome Res.*, 2, 495-505.
- ⁷⁷ Noguchi T., Hwang D.F., Arakawa O., Sugita H., Deguchi Y., Shida Y, Hashimoto K., **1987**, *Vibrio alginolyticus*, A Tetrodotoxin-producing Bacterium, in the Intestines of the Fish *Fugu vermicularis vermicularis*, *Mar. Biol.*, 94, 625-630.
- ⁷⁸ Daly G.W., Ernest Guenther award in chemistry of natural products, Amphibian skin: a remarkable source of biological active arthropods alkaloids, **2003**, *J. Med. Chem.*, 46, 445-452.
- ⁷⁹ Badio B., Daly J.W., Epibatidine, a potent analgesic and nicotinic agonist, **1994**, *Mol. Pharmacol.*, 45, 563-569.
- ⁸⁰ Li X., Eisenach J.C., Nicotinic acetylcholine receptor regulation in the spinal norepinephrine release, **2002**, *Anaesthesiology*, 96, 1450-1456.
- ⁸¹ Lewis R.J., Garcia M.L., **2003**, Therapeutic potential of venom peptides, *Nat. rev. Drug Discov.*, 2, 790-802.
- ⁸² Fox, J.W., Shannon J.D., Stefansson B., Kamiguti A.S., Theakston R.D.G., Serrano S.M.T., Camargo A.C.M., Sherman N., Role of discovery science in toxinology: examples in venom proteomics, **2002**, Dans “*Perspectives in Molecular Toxinology*”, Menez A. (Ed.), Wiley, 95-106.
- ⁸³ Stöcklin, R., Favreau P., Proteomics of Venom Peptides, **2002**, Dans “*Perspectives in Molecular Toxinology*”, Menez A. (Ed.), Wiley, 107-124.
- ⁸⁴ Malany S., Osaka H., Sine S.M., Taylor P., Orientation of the α -neurotoxin at the subunit interfaces of the nicotinic acetylcholine receptor, **2000**, *Biochem.*, 39, 15388-15389.
- ⁸⁵ Miyoshi S, Tu A.T., Muscarinic acetylcholine receptor (mAChR) inhibitor from snake venom: Interaction with subtypes of human mAChR, **1999**, *Ach. Biochem. Biophys.*, 369, 114-118.
- ⁸⁶ Freedman J.E., Snyder S.H., Vipoxin. A protein from Russell’s viper venom with high affinity for biogenic amine receptors, **1981**, *J. Biol. Chem.*, 256, 13172-13179.
- ⁸⁷ Yates S.L., Rosenberg P., Comparative effects of phospholipase A2 neurotoxins and enzymes on membrane potential and Na⁺/K⁺ ATPase activity of rat brain synaptosomes, **1991**, *Toxicol Appl Pharmacol*, 109, 207-218.
- ⁸⁸ Harvey A.L., Twenty years of dendrotoxins, **2001**, *Toxicon*, 39, 15-26.
- ⁸⁹ Karlsson E., Mbugua P.M., Rodriguez-Ithurralde D., Fasciculins, anticholinesterase toxins from the venom of the green mamaba *Dendroaspis angusticeps*, **1984**, *J. Physiol. (Paris)*, 79, 232-240.
- ⁹⁰ Bober M.A., Glenn J.L., Straight R.C., Ownby C.L., Detection of myotoxin alpha-like proteins in various snake venoms, **1988**, *Toxicon*, 26, 665-673.
- ⁹¹ Huang T.F., What have snakes taught us about integrins?, **1998**, *Cell. Mol. Life Sci.*, 54, 527-540.
- ⁹² Chippaux J.-P., Venins de serpent et envenimations, **2002**, Ed. IRD Editions, *Institut de Recherche pour le Développement*, Coll. Actiques, Paris.
- ⁹³ Hogue Angeletti R., Nerve growth factor in cobra venoms, **1970**, *Proc. Natl. Acad. Sci.*, 65, 668-675.
- ⁹⁴ Marsh N.A., Diagnostics uses of snake venoms, **2001**, *Haemostasis*, 31, 211-217.

- ⁹⁵ Beviglia L., Stewart G.L., Niewiarowski S., Effect of four disintegrins on the adhesive and metastatic properties of B16F10 melanoma cells in a murine model, *Oncol. Res.*, **1995**, 7, 7-20.
- ⁹⁶ Zhou Q., Nakada M.T., Brooks P.C., Swenson S.D., Ritter M.R., Argounova S., Arnold C., Markland F.S., Contortrostatin, a homodimeric disintegrin binds to integrin $\alpha v \beta 5$, *Biochem. Biophys. Res. Commun.*, **2000**, 267, 350-355.
- ⁹⁷ Ondetti M.A., Williams N.J., Sabo E.F., Pluscec J., Weaver E.R., Kocy O., Angiotensin-converting enzyme inhibitors from the venom of *Bothrops jararaca*. Isolation, elucidation of structure and synthesis, 1971, *Biochemistry*, **19**, 4033-4039.
- ⁹⁸ Favreau P., Le Gall F., Molgo J., Le venin des cônes : source de nouveaux outils pour l'étude de récepteurs et canaux ioniques, *Annales de l'Institut Pasteur*, **1999**, 10, 273-284.
- ⁹⁹ Olivera B.M., Conus Venom Peptides, Receptors and Ion Channel Targets, and Drug Design : 50 Millions Years of Neuropharmacology, *Mol. Biol. Cell*, **1997**, 8, 2101-2109.
- ¹⁰⁰ Olivera B.M., Rivier J., Clark C., Ramilo C.A., Corpuz G.P., Abogadie F.C., Mena E.E., Woodward S.R., Hillyard D.R., Cruz L.J., Diversity of conus neuropeptides, **1990**, *Science*, 249, 257-263.
- ¹⁰¹ McIntosh J.M., Olivera B.M., Gray W.R., γ -carboxyglutamate in a neuractive toxin, **1984**, *J. Biol. Chem.*, 259, 14343-14346.
- ¹⁰² Grey Craig A., Badyopadhyay P., Olivera B.M., Post-translationally modified neuropeptides from *Conus* venoms, **1999**, *Eur. J. Biochem.*, 264, 271-275.
- ¹⁰³ Myers R.A., McIntosh J.M., Imperial J., William R.W., Oas T., Haack J.A., Hernandez J.F., Rivier J., Cruz L.J., Olivera B.M., Peptides from *Conus* venoms which affect Ca^{++} entry into neurons, **1990**, *J. Toxicol. Toxin Rev.*, 9, 179-2002.
- ¹⁰⁴ Terlau H., Olivera B.M., *Conus* Venoms: a Rich Source of Ion Channel-Targeted Peptides, **2004**, *Physiol. Rev.*, 84, 41-68.
- ¹⁰⁵ Hui K., Lipkind G., Fozzard H.A., French R.J., Electrostatic and steric contributions to block of the skeletal muscle sodium channel by μ -conotoxin, **2002**, *J. gen. Physiol.*, 119, 45-54.
- ¹⁰⁶ Terlau H., Stocker M., Shon K.J., McIntosh J.M., Olivera B.M., μ O-conotoxin, MrVIA inhibits mammalian sodium channels but not through site I, **1996**, *J. Neurosci.*, 16, 1423-1429.
- ¹⁰⁷ Hille B., *Ion Channels of Excitable Membranes* (3rd ed.), 2001, Sunderland, MA: Sinauer.
- ¹⁰⁸ Shon K.J., Stockers M., Terlau H., Stühner W., Jacobsen R., Walker C., Grilley M., Watkins M., Hillyard D.R., Gray W.R., Olivera B.M., k-conotoxin PVIIA is a peptide inhibiting the Shaker K⁺ channel, **1998**, *J. Biol. Chem.*, 273, 33-38.
- ¹⁰⁹ Ferber M., Sporning A., Jeserich G., DeLaCruz R., Watkins M., Olivera B.M., Terlau H., A Novel *Conus* Peptide Ligand for K⁺ Channels, **2003**, *J. Biol. Chem.*, 278, 2177-2183.
- ¹¹⁰ Colledge C.J., Hunsperger J.P., Imperial J.S., Hillyard D.R., Precursor structure of omega-conotoxin GVIA determined from a cDNA clone, **1992**, *Toxicon*, 30, 1111-1116.
- ¹¹¹ Olivera B.M., Cruz L.J., de Santos V., Neuronal calcium channel antagonists. Discrimination between calcium channel subtypes using omega-conotoxin from *Conus magus* venom, **1987**, *Biochem.*, 26, 2086-2090.
- ¹¹² McIntosh J.M., Santos A.D., Olivera B.M., *Conus* peptides targeted to specific nicotinic acetylcholine receptor subtypes, 1999, *Annu. Rev. Biochem.*, **68**, 59-88.

- ¹¹³ England L.J., Imperial J., Jacobsen R., Grey Craig A., Gulyas J., Akhtar M., Rivier J., Julius J., Olivera B.M., Inactivation of a Serotonin-Gated Ion Channel by a Polypeptide Toxin from Marine Snails, **1998**, *Science*, 281, 575-578.
- ¹¹⁴ Alonso D., Khalil Z., Satkunanthan, Livett B.G., Drugs from the Sea: Conotoxins as Drugs for Neuropathic Pain and Other Neurological Conditions, **2003**, *Mini Reviews in Medical Chemistry*, 3, 785-787.
- ¹¹⁵ Li S., Wang J., Zhang X., Ren Y., Wang N., Zhao K., Chen X., Zhao C., Li X., Shao J., Yin J., West M.B., Xu N., Liu S., Proteomic characterization of two snake venoms: *Naja naja atra* and *Agkistrodon halys*, **2004**, *Biochem.*, 384, 119-127.
- ¹¹⁶ Legros C., Celerier M.L., Henry M., Guette C., Nanospray analysis of the venom of the tarantula *Theraphosa leblondii*: a powerful method for direct venom mass fingerprinting and toxin sequencing, 2004, *Rapid Commun. Mass Spectrom.*, 18, 1024-1032.
- ¹¹⁷ Escoubas P., Chamot-Rooke J., Stocklin R., Whiteley B.J., Corzo G., Genet R., Nakajima T., A comparison of matrix-assisted laser desorption-ionization time-of-flight and liquid chromatography electrospray mass spectrometry for the identification of tarantula venoms samples, **1999**, *Rapid Commun. Mass Spectrom.*, 12, 1075-1084.
- ¹¹⁸ Jones A., Bingham J.P., Gehrmann J., Bond T., Loughnan M., Atkins A., Lewis R.J., Alewood P.F., Isolation and characterization of conopeptides by high-performance liquid chromatography combined with mass spectrometry and tandem mass spectrometry, **1996**, *Rapid Commun. Mass Spectrom.*, 10, 138-143.
- ¹¹⁹ Bazaa A., Marrakchi N., El Ayebe M., Sanz L., Calvete J.J., Snake venomomics: comparative of the venom proteomes of the Tunisian snakes *Cerastes cerastes*, *Cerastes vipera* and *Macrovipera lebetina*, **2005**, *Proteomics*, 16, 4223-4235.
- ¹²⁰ Favreau P., Menin L., Michalet S., Perret F., Cheneval O., Stocklin M., Bulet P., Stocklin R., Mass spectrometry strategies for venom mapping and peptide sequencing from crude venoms : Case applications with single arthropod specimen, **2006**, *Toxicon*, 47, 676-687.
- ¹²¹ Cain T.C., Lubman D.M., Weber W.J.J. Differentiation of bacteria using protein profiles from matrix-assisted laser desorption/ionization mass spectrometry, **1994**, *Rapid Commun. Mass Spectrom.*, 8, 1026-1030.
- ¹²² Fry B.J., Wickramaratna J.C., Hodgson W.C., Alewood P.F., Kini R.M., Ho H., Wuster W., Electrospray liquid chromatography/mass spectrometry fingerprinting of *Acanthophis* (death adder) venoms: taxonomic and toxinological implications, **2002**, *Rapid Commun. Mass Spectrom.*, 16, 600-608.
- ¹²³ Romy-Lebrun R., Martin-Eauclaire M.F., Escoubas P., Wu F.Q., Lebrun B., Hisada M., Nakajima T., Characterization of four toxins from *Butus martensi* scorpion venom which acts on apamin-sensitive Ca²⁺-activated K⁺ channels, **1997**, *Eur. J. Biochem.*, 245, 457-464.
- ¹²⁴ Wilson D., Alewood P.F., Taxonomy of Australian Funnel-web spiders using rp-HPLC/ESI-MS profiling techniques, **2006**, *Toxicon*, 47, 614-627.
- ¹²⁵ Escoubas P., Celerier M.L., Nakajima T., High-performance liquid chromatography matrix-assisted laser desorption/ionization time-of-flight mass spectrometry peptide fingerprinting of tarantula venoms in the genus *Brachypelma*: chemotaxonomic and biochemical application, **1997**, *Rapid Commun. Mass Spectrom.*, 11, 1891-1899.
- ¹²⁶ Escoubas P., Sollod B., King G.F. Venom landscapes : mining the complexity of spider venoms via a combined cDNA and mass spectrometric approach, **2006**, *Toxicon*, 47, 650-663.

- ¹²⁷ Menez A., Stöcklin R., Mebs D., „Venomics“ or the venomous systems genome project, **2006**, *Toxicon*, 47, 255-259.
- ¹²⁸ Guijarro J.I., M'Barek S., Gomez-Lagunas F., Garnier D., Rochat H., Sabatier J.M., Possani L., Delepierre M., Solution structure of Pi4, a short four-disulfide-bridged scorpion toxin specific of potassium channels, **2003**, *Protein Sci.*, 12, 1844-1854.
- ¹²⁹ Moller C., Rahmankhah S., Lauer-Fields J., Bubis J., Fields G.B., Mari F., A novel conotoxin framework with a helix-loop-Helix (Cs α/α) Fold, **2005**, *Biochem.*, 44, 15986-15996.
- ¹³⁰ Louhnan M., Nicke A., Jones A., Schroeder C.I., Nevin S.T., Adams D.J., Alewood P.F., Lewis R.J., Identification of a novel class of nicotinic receptor antagonists: dimeric conotoxins VxIIA, VxXIIB and VxXIIC from *Conus vexillum*, **2006**, *J. Biol. Chem.*, *in press*.
- ¹³¹ Fox J.W., Ma L., Nelson K., Sherman N.E., Serrano S.M., Comparison of indirect and direct approaches using ion-trap and Fourier transform ion cyclotron resonance mass spectrometry for exploring viperid venom proteomes, **2006**, *Toxicon*, 47, 700-714.

Chapitre I

Caractérisation de nouvelles sarafotoxines dans un
venin brut de serpent par spectrométrie de masse
FT-ICR et clonage moléculaire

I. Introduction

Les sarafotoxines (SRTXs) forment avec les endothélines (ETs) une famille homogène de puissants isopeptides vasoconstricteurs constitués d'une vingtaine d'acides aminés et de deux ponts disulfures intramoléculaires^{1,2}. SRTXs et ETs agissent au niveau du système vasculaire en affectant la contraction cardiaque ainsi que les muscles lisses dans différents tissus chez les vertébrés *via* des récepteurs identiques. Surproduites de manière spécifique dans les venins des serpents du genre *Atractaspis* au sein de la famille des *Atractaspididae*, les SRTXs font partie des toxines de serpent les plus toxiques décrites à ce jour, capables d'entraîner la mort en quelques minutes chez la souris et moins d'une heure chez l'homme. Les ETs quant à elles, sont produites en très faibles quantités par les cellules endothéliales du système cardiovasculaire des vertébrés et ont un mode d'action de type hormonal. Elles semblent principalement impliquées dans la régulation du système vasculaire, ainsi que dans diverses manifestations cardio-pathologiques. Les similitudes structurales et pharmacologiques qui existent entre les SRTXs et les ETs suggèrent fortement que ces dernières soient des équivalents endogènes des toxines chez les vertébrés ce qui rend l'étude des SRTXs particulièrement attrayante.

Les sarafotoxines ont été caractérisées pour la première fois dans le venin d'un serpent de l'espèce *Atractaspis engaddensis*⁵. Jusqu'à récemment, les sarafotoxines caractérisées étaient toutes constituées de 21 acides aminés et 2 ponts disulfures. Une analyse du venin d'*Atractaspis microlepidota microlepidota* a révélé très récemment (2004) l'existence de six nouvelles isoformes de SRTXs plus longues de trois acides aminés du côté de l'extrémité C-terminale³. Par ailleurs, deux études de la structure des précurseurs codant les sarafotoxines d'*Atractaspis engaddensis* et *Atractaspis microlepidota microlepidota* ont montré de manière originale et inattendue que leur organisation présente une structure répétitive constituée d'une séquence de SRTX précédée par un peptide « espaceur » invariant. Cette organisation dite en « chapelet » constitue une stratégie simple, économique et efficace d'amplification de la production de SRTXs. Elle est unique par sa régularité au sein des précurseurs codant des toxines de serpent à ponts disulfures. Qu'il s'agisse en effet de toxines dites « à trois doigts » telles les neurotoxines curarisantes, les toxines muscariniques, les fasciculines, ou de phospholipases de type A₂ pour lesquelles de nombreuses isoformes ont été identifiées, toutes sont produites à partir de précurseurs codant invariablement une seule isoforme.

Les études réalisées jusqu'à présent l'ont été par une approche classique : séparation des différents constituants du venin par plusieurs étapes de chromatographie liquide, caractérisation des toxines isolées par dégradation d'Edman et spectrométrie de masse en tandem. Dans ce contexte, nous avons souhaité évaluer l'intérêt d'utiliser la spectrométrie de masse FT-ICR pour l'étude d'un échantillon unique du venin d'*Atractaspis irregularis*, jamais étudié au niveau moléculaire et susceptible de contenir de nouvelles sarafotoxines. Une étude par clonage moléculaire des ARNm des glandes à venin a été réalisée en parallèle par nos collaborateurs biologistes du CEA (Dr. F. Ducancel) afin de déterminer la structure d'éventuels précurseurs de sarafotoxines.

La première partie de ce chapitre sera consacrée à un rappel bibliographique concernant les sarafotoxines et les endothélines : leurs structures respectives et leur mode d'action. Le protocole d'analyse par spectrométrie de masse FT-ICR, mis au point pour réaliser une cartographie du venin brut d'*Atractaspis irregularis* sera décrit dans une deuxième partie. La description des deux approches choisies pour caractériser précisément de nouvelles sarafotoxines dans le venin brut (clonage moléculaire et séquençage *de novo* par spectrométrie de masse) constitue la troisième et dernière partie de ce chapitre.

II. Les sarafotoxines de serpents: des analogues structuraux et fonctionnels des endothélines des mammifères

A. Découverte et historique des sarafotoxines

Les sarafotoxines sont des isopeptides que l'on trouve uniquement dans les venins des serpents de l'espèce *Atractaspis*. Les *Atractaspis* sont de petits serpents fouisseurs de 30 à 70 centimètres de long qui vivent sous terre et qui se nourrissent principalement de petits reptiles, d'amphibiens et de rongeurs. Ils sont localisés essentiellement en Afrique sub-saharienne mais sont également représentés à un degré moindre en Israël et dans le sud-ouest de la péninsule arabique. L'une des particularités de ces serpents est la manière atypique d'envenimation qu'ils utilisent. Ils ont la possibilité de conserver l'un des deux crochets à venin sorti tout en maintenant la bouche fermée, ce qui les rend capables de piquer leurs proies la bouche fermée en se servant du crochet comme d'un véritable poignard (*Figure 1*). Cette observation est unique chez les reptiles.

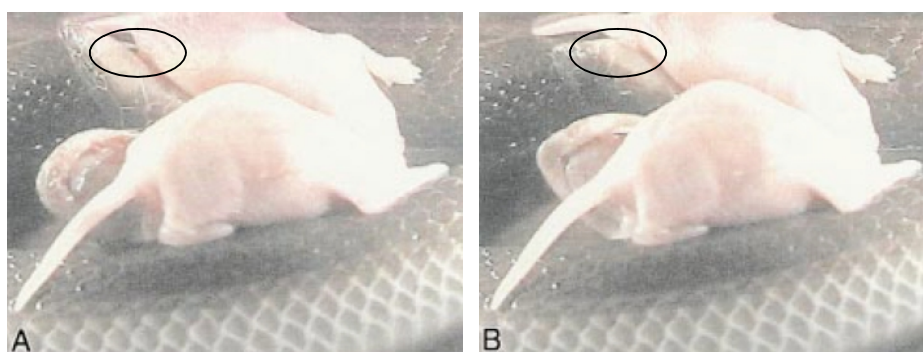


Figure 1: Envenimation d'une souris par un serpent *Atractaspis bibroni*⁴

(A) Le crochet est sorti de la bouche bien que celle-ci soit fermée.

(B) le crochet pénètre dans la proie par un vif mouvement de la tête du serpent.

La forme de la tête des *Atractaspis* montre également que les crochets à venin ne servent pas uniquement à l'envenimation mais aussi à la manipulation et au transport des proies⁴.

Le mot « sarafotoxine » (SRTX) dérive du nom hébreu du serpent *Atractaspis engaddensis* (*Saraf Ein Gedi*) dans le venin duquel l'équipe d'E. Kochva les a découvertes pour la première fois en 1982^{5,6}. La nomenclature adoptée pour les trois premières sarafotoxines décrites est : SRTX-a, -b et -c⁷. Ces sarafotoxines présentent, entre elles, une homologie de séquence d'au moins 85% et possèdent une extrémité C-terminale totalement conservée. Une dizaine d'années plus tard, en 1993, une nouvelle sarafotoxine, ne différant de la SRTX-b que par la substitution d'une lysine par une alanine en position 4 est découverte dans le venin d'*Atractaspis bibroni*, donnant le nom de bibrotoxine à ce nouveau peptide vasoconstricteur⁸. La même année, le premier clonage moléculaire d'un précurseur des SRTXs d'*Atractaspis engaddensis* permet d'identifier 3 nouvelles isoformes nommées a1, b1 et e⁹. Plus récemment, en 2004, une famille de sarafotoxines atypique est découverte dans le venin du serpent *Atractaspis microlepidota microlepidota*³. Jusqu'à cette étude, toutes les sarafotoxines possédaient des séquences de 21 acides aminés. Cette nouvelle famille, nommée SRTX-m_i, est constituée de toxines plus longues de 3 acides aminés du côté de l'extrémité C-terminale, offrant ainsi une diversité supplémentaire pour ce type de composés.

Du point de vue de leur activité biologique, les sarafotoxines font partie des toxines de serpents les plus toxiques décrites à ce jour. La dose létale moyenne (DL₅₀) est d'environ 15 µg/kg¹⁰, les plus toxiques étant les SRTX-a et -b dont la DL₅₀ est évaluée à 10 µg/kg.

B. Les endothélines des mammifères : rôle et mode d'action

Les endothélines sont des petits peptides de 21 acides aminés essentiellement fabriqués par les cellules endothéliales. Elles ont une activité vasoconstrictrice et modulent notamment la contraction du cœur et des muscles lisses de différents tissus. Grâce à cette action, les endothélines sont impliquées non seulement dans la régulation du système vasculaire mais également dans le développement de plusieurs maladies cardio-vasculaires. A titre d'exemple, une augmentation des endothélines plasmatiques peut être observée au cours des hypertension artérielles, l'infarctus du myocarde, l'insuffisance cardiaque et la maladie de Raynaud qui se traduit par un trouble de la circulation au niveau des petites artères des extrémités (doigts, orteils, nez, oreilles). Les endothélines sont produites essentiellement par l'endothélium vasculaire¹¹ mais elles sont présentes également dans diverses glandes endocrines comme les gonades et la glande corticosurrénale.

Le système endothélial des mammifères se réduit à trois peptides nommés ET-1, ET-2 et ET-3. Un quatrième, appelé VIC pour *vasoactive intestinal contractor* a été caractérisé uniquement chez les rongeurs¹². Classiquement, ces peptides sont composés de 21 acides aminés et possèdent deux ponts disulfures conservés entre les cystéines +1/+15 et +3/+11, constituant une signature unique de cette famille. Récemment, de nouvelles formes bioactives d'endothélines ont été caractérisées¹³ montrant une séquence analogue aux ET-1, -2 et -3 mais qui se prolonge de 10 acides en aminés en position C-terminale (*Figure 2*).

Mammifères	ET-1	C	S	C	S	S	L	M	D	K	E	C	V	Y	F	C	H	L	D	I	I	W																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									
------------	------	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Figure 2: Séquences en acides aminés des endothélines présentes chez les mammifères

Leur biosynthèse s'effectue à partir de prépropeptides. L'ET-1 humaine est d'abord synthétisée en un précurseur de 212 acides aminés nommé Pre-pro-ET-1 (*Figure 3*). La séquence de l'endothéline finale est localisable entre la cystéine 53 et le tryptophane 73. Ce précurseur subit alors l'action d'endopeptidases spécifiques des doublets basiques. Les ruptures occasionnées par cette enzyme entraînent la formation d'un nouveau peptide de 39 acides aminés nommé Big-ET-1. Ce gros peptide, sous l'influence de l'enzyme de conversion

des endothélines (enzyme à zinc), conduit à la forme active de 21 acides aminés de l'endothéline, ET-1¹⁴.

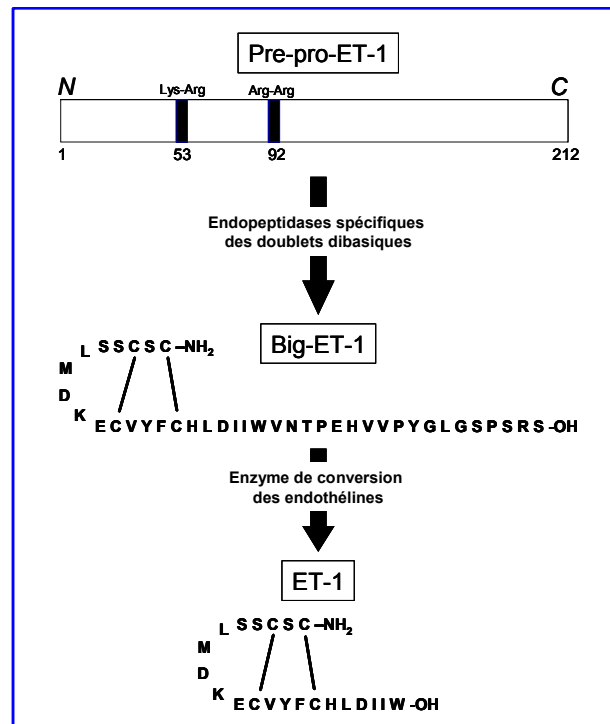


Figure 3 : Voie de synthèse de l'endothéline ET-1

Les endothélines libérées peuvent activer deux types de récepteurs : le type A (ET_A) et le type B (ET_B). ET_A et ET_B sont localisés sur les fibres vasculaires lisses mais les ET_B sont également présents au niveau des cellules endothéliales. Lors de l'activation par les endothélines, il y a vasoconstriction des fibres vasculaires et on observe une libération de monoxyde d'azote et de prostacycline au niveau des cellules endothéliales. La libération de ces composés entraîne notamment une vasodilatation. La vasoconstriction occasionnée est intense et longue, ce qui contraste avec le temps de demi-vie des endothélines, de l'ordre de la minute.

C. Comparaison sarafotoxines/endothélines

C.1. Homologies structurales

De nombreuses similitudes entre les sarafotoxines et les endothélines ont été mises en évidence. La première concerne la séquence primaire des deux familles. Leur comparaison

(Figure 4) montre que ces séquences possèdent environ 60% d'homologies. Plus précisément, il est intéressant de remarquer que l'extrémité C-terminale des peptides est plus conservée que l'extrémité N-terminale.

[illegible]

Figure 4 : Alignement des séquences des sarafotoxines connues et comparaison avec celles des endothélines humaines

Différentes techniques (diffraction des rayons X¹⁵, résonance magnétique nucléaire¹⁶ et modélisation moléculaire¹⁷) ont été utilisées pour déterminer la structure tridimensionnelle des endothélines. Des études par RMN ont également été menées pour la sarafotoxine SRTX-b¹⁸. De très grandes similitudes sont alors apparues entre les structures des deux familles. Sarafotoxines et endothélines présentent une structure étendue au niveau de l'extrémité *N*-terminale des acides aminés +1 à +4, puis un agencement coudé entre les +5 et +8 suivi d'une hélice α entre les acides aminés +9 et +15 stabilisée par les deux ponts disulfures. (*Figure 5*). En revanche, l'extrémité *C*-terminale ne semble pas adopter de conformation particulière.

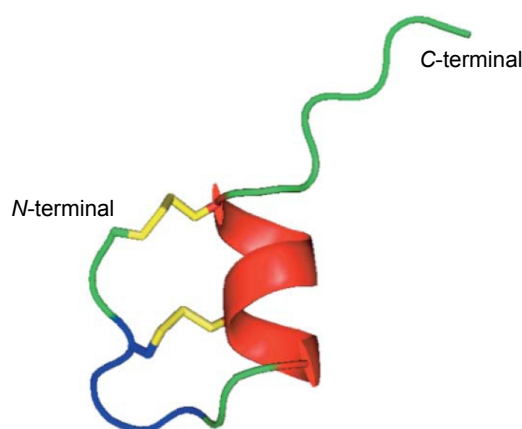


Figure 5 : Structure tridimensionnelle de SRTX-b obtenue par RMN

Profitant du fait que la tyrosine en position +13 peut être marquée à l'iode 125, des expériences de liaison (« binding ») ont été effectuées sur la SRTX-b. Ces expériences ont pour but de déterminer l'activité biologique de cette toxine et plus particulièrement de cibler les récepteurs auxquels elle se lie. Il a ainsi été démontré que la SRTX-b iodée lie spécifiquement et avec une très bonne affinité les préparations de membranes de cellules cardiaques atriales⁷. Cette liaison peut être inhibée par les sarafotoxines SRTX-a et SRTX-c et par les endothélines de 21 acides aminés, ce qui suggère fortement qu'elles possèdent le même site de liaison. Par ailleurs, différents ligands connus de récepteurs membranaires et de canaux ioniques n'inhibent pas la liaison SRTX-récepteur suggérant que les sarafotoxines se lient à de nouvelles cibles moléculaires.

C.2. Organisation des précurseurs

Alors que les endothélines sont produites en très faibles quantités chez les mammifères (concentration picomolaire dans les plasmas), les SRTXs sont toujours surproduites dans les venins des *Atractaspidae*. A partir de cette observation et avec l'objectif d'étudier le métabolisme des sarafotoxines dans les glandes à venin, une étude, par biologie moléculaire, de la structure des précurseurs de plusieurs sarafotoxines issues de venins différents, a été entreprise.

Dans un premier temps une banque d'ADN complémentaire (ADNc) a été construite à partir des ARN messagers (ARNm) purifiés à partir des glandes à venin d'*A. engaddensis*⁹. Une banque d'environ 7000 clones a été obtenue avant d'être criblée par une séquence nucléotidique correspondant aux 7 derniers acides aminés très conservés des SRTXs. Quatorze clones sont alors isolés, dont la taille maximale observée est de 2000 paires de bases. La *Figure 6* représente la séquence nucléotidique d'un des ADNc obtenus. Les nucléotides représentés en italique correspondent à l'extrémité non codante du précurseur. La correspondance de cette séquence en acides aminés est donnée pour les codons situés entre la méthionine d'initiation et le codon final « TAG », aussi couramment appelé «codon stop ». La méthionine est suivie d'un peptide hydrophobe caractéristique d'une séquence signal. L'originalité de ce précurseur est la présence d'une structure répétitive de douze séquences de 40 acides aminés (39 pour la première séquence).

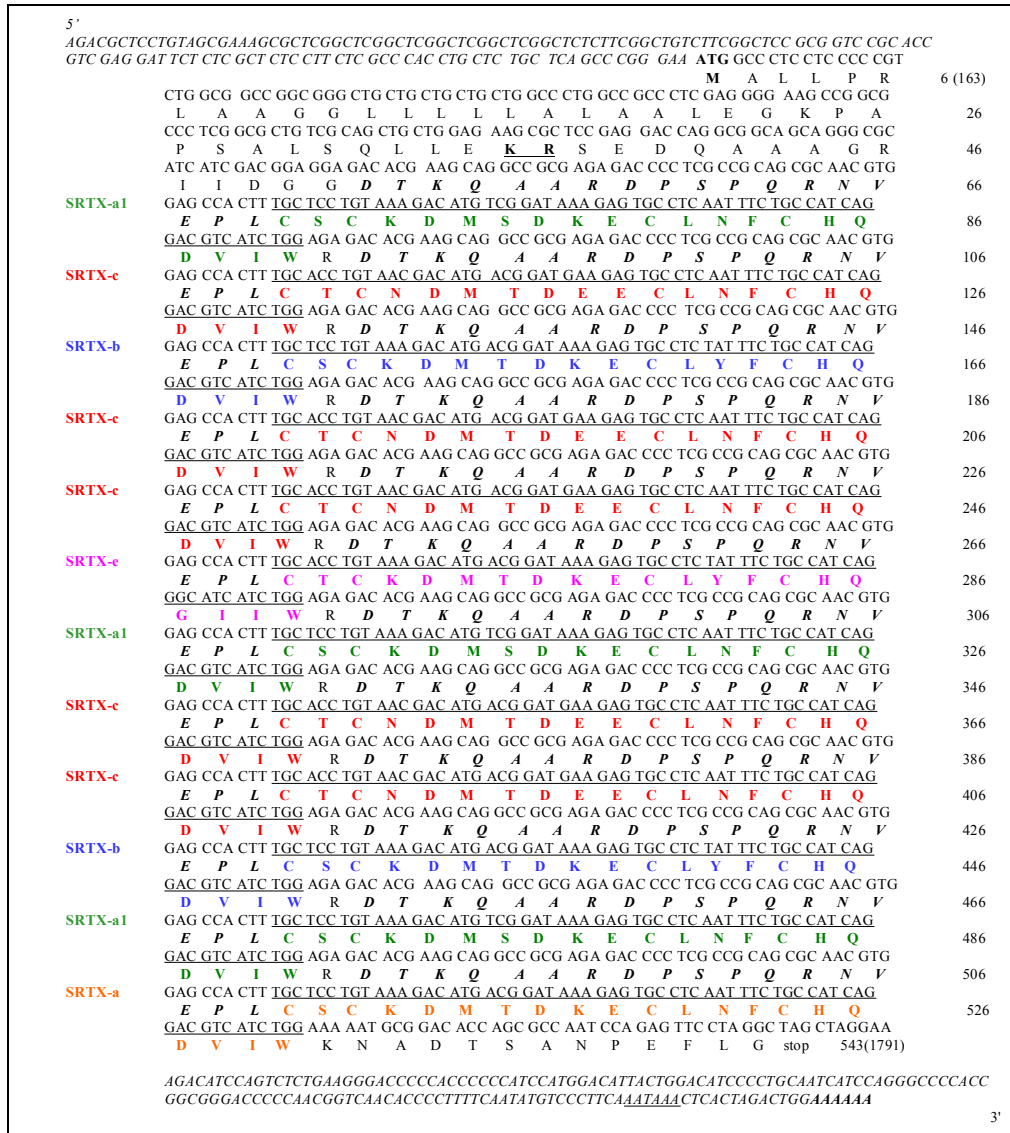


Figure 6: Structure complète du précurseur codant les SRTXs du serpent *Atractaspis Engaddensis*

Après traduction de la séquence nucléotidique en acides aminés, il apparaît que chacune des 12 structures répétées contient une séquence de SRTX. Ces séquences de sarafotoxines sont séparées dans le précurseur par un peptide espaceur invariant de 19 acides aminés (18 dans le premier cas). En tout, cinq isoformes de sarafotoxines différentes sont identifiées. On remarquera que leur séquence est très conservée puisque seulement six acides aminés varient. Une sixième isoforme appelée SRTX-b1 qui ne diffère de la SRTX-b que par la présence d'une sérine en position +7 à la place de la thréonine (différence identique à celle présente entre SRTX-a et SRTX-a1) a été identifiée à partir d'un autre ADN⁹. Six sarafotoxines différentes ont donc été identifiées à partir de l'analyse des ARN messagers des glandes à venin d'*Atractaspis engaddensis*. Cette organisation polycistronique du précurseur est très originale en ce qui concerne les précurseurs des toxines animales possédant deux ponts

disulfures. A titre d'exemple, les toxines dites à trois doigts (toxines muscariniques, neuromusculaires, fasciculines, les phospholipases A₂) sont toutes produites par l'intermédiaire d'un précurseur monocistronique⁹. Ce qui est également remarquable est que la SRTX-c, l'isoforme la plus représentée au sein du précurseur (5 répliques), est la toxine la plus concentrée dans le venin. Ce type d'organisation s'apparente donc à un moyen fiable et économique de surproduction des sarafotoxines au sein du venin d'*Atractaspis engaddensis*. Par ailleurs, il est à noter que cette structure est très différente de celle des précurseurs des endothélines¹², ces dernières étant produites à partir de précurseurs monocistroniques. Une étude similaire a été menée dans un deuxième temps sur le serpent *Atractaspis microlepidota microlepidota*³. Dans cette étude, la banque d'ADNc a été construite par RT-PCR via l'utilisation d'amorces nucléotidiques complémentaires des extrémités 5' et 3' non codantes des précurseurs séquencés chez *A. engaddensis*. La RT-PCR puis le clonage plasmidique ont permis d'amplifier uniquement les ADNc susceptibles de coder pour des précurseurs de sarafotoxines chez *A. m. microlepidota*.

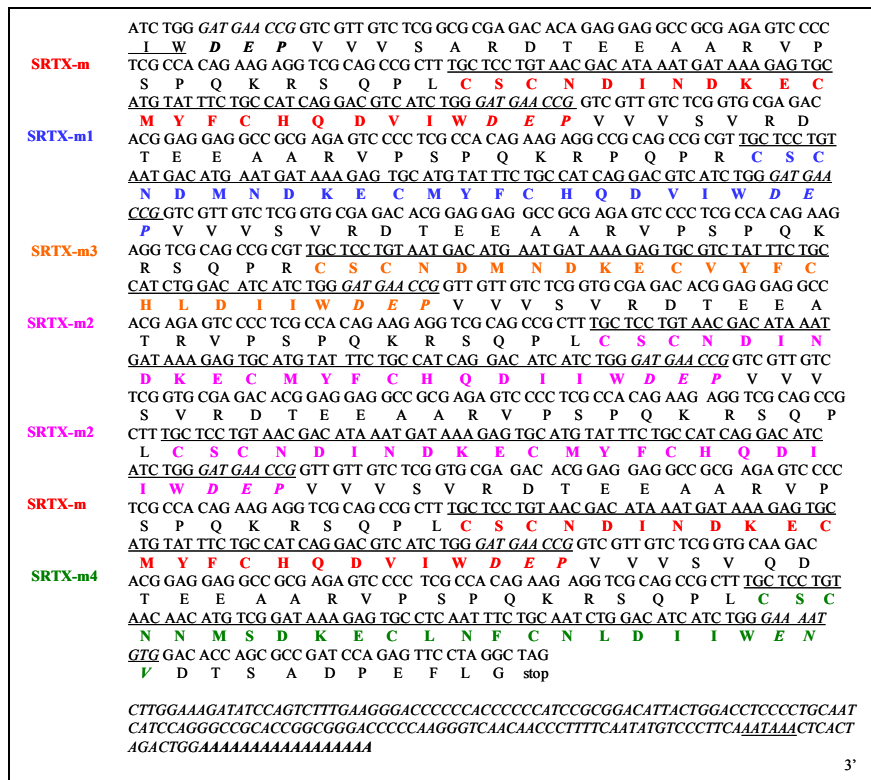


Figure 7: Structure complète du précurseur codant les SRTXs du serpent *Atractaspis microlepidota microlepidota*

Cette étude a également mis en évidence le caractère polycistronique du précurseur codant pour les sarafotoxines. La séquence du précurseur obtenu est détaillée dans la Figure 7.

Même s'il est incomplet, le précurseur montre clairement que son organisation polycistronique s'apparente à celle du précurseur décrit précédemment, sans toutefois être identique. En effet, le motif répétitif est constitué cette fois de 48 acides aminés (répétition de 144 nucléotides) contenant une séquence de sarafotoxine suivie d'un peptide espaceur de 24 acides aminés. Cinq isoformes ont été découvertes, SRTX-m, -m1, -m2, -m3 et -m4. La principale différence entre cette étude et celle menée sur *A. engaddensis* est que les isoformes trouvées sont constituées de 24 acides aminés, et que les 3 acides aminés supplémentaires sont situés du côté de l'extrémité C-terminale. Ces trois acides aminés supplémentaires sont : un acide aspartique (D), un acide glutamique (E) et une proline (P), pour les SRTX-m, -m1, -m2 et -m3 et un acide glutamique (E), une asparagine (N), et une valine (V) pour la SRTX-m4. Il est également important de remarquer que des formes de sarafotoxines plus courtes ont été détectées dans le venin. Deux principales hypothèses peuvent expliquer la présence de ces composés. Tout d'abord, ces formes peuvent résulter de la dégradation des sarafotoxines plus longues par l'action d'enzymes comme des endopeptidases. Ensuite, chacune d'entre elles peut posséder son propre rôle biologique. Seule une étude de leur activité biologique permettrait de conclure. La SRTX-m a été synthétisée chimiquement puis étudiée plus en détail par RMN. L'objectif était d'évaluer l'influence des trois acides aminés supplémentaires sur la structure tridimensionnelle de la toxine. Les résultats montrent que la présence de ces acides aminés supplémentaires n'induit pas de modification notable de la structure tridimensionnelle décrite pour les sarafotoxines à 21 acides aminés. Notamment, aucune structuration supplémentaire au niveau de l'extrémité C-terminale n'apparaît. Par ailleurs, le degré de toxicité chez la souris (essentiellement par vasoconstriction) de la SRTX-m est un peu plus faible que celui de la SRTX-b. De plus, bien que la SRTX-m et la SRTX-b induisent des symptômes identiques, lorsqu'elles sont injectées à une souris, des expériences de liaison montrent que la fixation de la SRTX-m n'entre pas en compétition avec celle de la SRTX-b marquée. Ceci suggère, soit que la SRTX-m ne reconnaît pas le même site de liaison sur le récepteur, soit qu'elle cible un nouveau sous-type de récepteur.

En conclusion, ces deux études ont montré que, chez les *Atractaspididae*, les précurseurs des sarafotoxines semblent adopter une organisation polycistronique. Les sarafotoxines décrites chez *A. engaddensis* comportent 21 acides aminés et deux ponts disulfures, comme les endothélines de mammifères. Il a été montré récemment que chez *A. microlepidota microlepidota*, les sarafotoxines pouvaient atteindre 24 acides aminés. Même si cette

élongation ne semble pas avoir de réelle signification biologique, il nous a semblé intéressant de poursuivre l'étude de ces toxines par l'analyse du venin d'un serpent de la même famille, *Atractaspis irregularis*, venin qui n'avait jamais été étudié au niveau moléculaire. Ce venin nous permettra tout d'abord d'étudier l'hétérogénéité des longueurs peptidiques des sarafotoxines au sein des venins de la famille des *Atractaspidae* mais également de montrer les intérêts que présente l'utilisation d'un spectromètre FT-ICR en toxinologie. Nous étudierons enfin la complémentarité de spectrométrie de masse FT-ICR avec des expériences de clonage moléculaire.

III. Cartographie du venin brut d'*Atractaspis irregularis* par FT-ICR

A. Analyse du venin brut à l'état natif par nano-electrospray-FT-ICR

La Figure 8 présente le spectre obtenu lors de l'analyse du venin brut par nanoESI-FTICR.

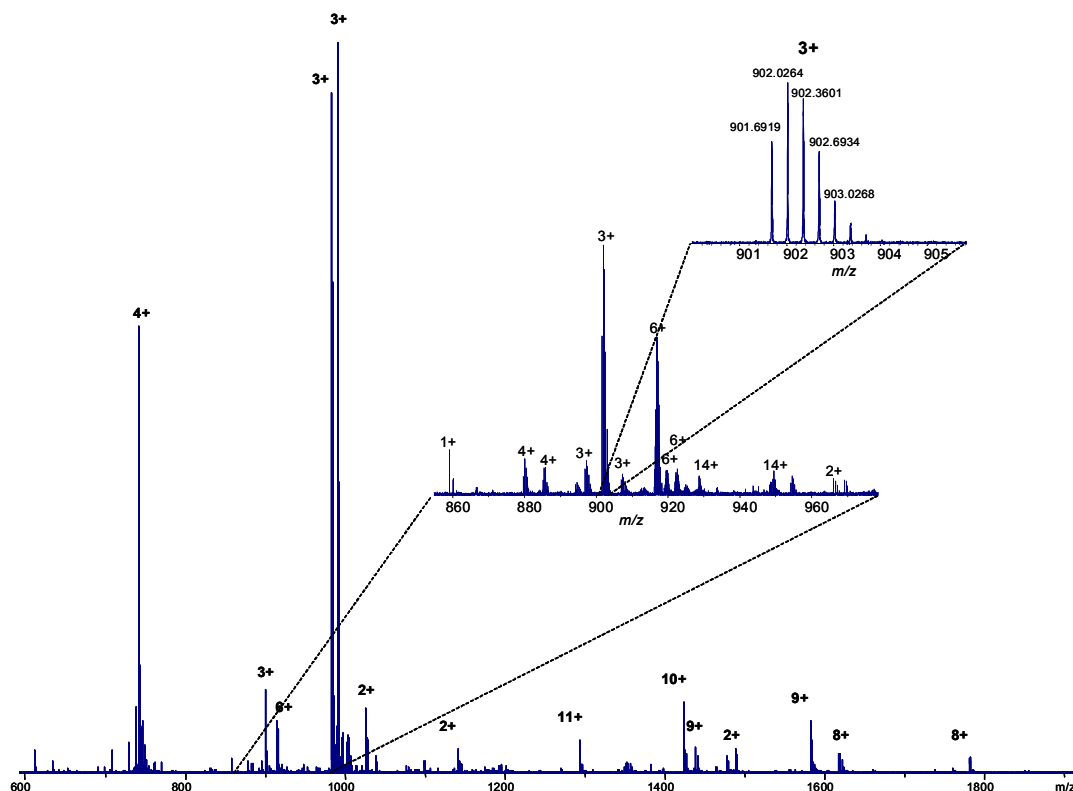


Figure 8: Spectre obtenu pour l'analyse du venin de *A. irregularis* par nanoESI-FTICR. Un zoom de région m/z [800 ; 980] a été ajouté pour montrer la complexité du spectre

L'analyse du spectre permet de dénombrer environ 80 composés différents compris dans une large gamme de masses moléculaires qui s'étend de 860 à 15 000 Da (*Tableau 1*). La richesse des espèces moléculaires présentes dans le venin est illustrée par l'agrandissement de la zone de m/z comprise entre 600 et 980.

Masses moléculaires en daltons			
614,3035	2456,9094	3018,2025	5946,3972
636,2991	2474,9249	3020,2137	6690,9553
655,3374	2680,0071	3034,1831	6961,9626
730,3340	2688,0291	3037,1837	6976,9661
858,3933	2702,0493	3085,2778	7613,3364
927,0800	2734,0428	3271,3175	8343,9918
999,1191	2841,1293	3293,3572	12917,4319
1002,2108	2859,1105	3353,4441	12950,4199
1026,4856	2926,2049	3515,4929	12981,9043
1419,6683	2933,1583	3537,5296	12983,2015
1666,4932	2940,2232	4606,8665	13053,7498
1929,8034	2950,1462	4633,8773	13274,3187
1935,6807	2958,1769	4703,9061	14213,0452
2027,7254	2966,1492	4730,9292	14215,2733
2050,7077	2972,1857	4774,9651	14230,4672
2066,7028	2982,1409	4801,9461	14243,3402
2076,6872	2988,1902	5392,3077	15605,2892
2175,7575	2994,2129	5421,1969	15808,3151
2270,8318	2998,2238	5490,5389	24848,1094
2280,8160	3002,1814	5506,7215	24851,1471
2288,8368	3004,1853	5524,4713	
2296,8174	3010,1361	5539,5025	

Tableau 1: Masses monoisotopiques mesurées lors de l'analyse du venin brut d'*A. irregularis*

Il est important de noter que la résolution apportée par la spectrométrie de masse FT-ICR (environ 80 000 à m/z 1000 dans notre étude) permet de déterminer les états de charge de chaque massif isotopique et de déduire la masse monoisotopique de chaque composé. La résolution apportée par d'autres instruments de type temps de vol et piège à ions n'aurait pas été suffisante pour résoudre tous les massifs isotopiques observés dans le spectre (*Figure 9*).

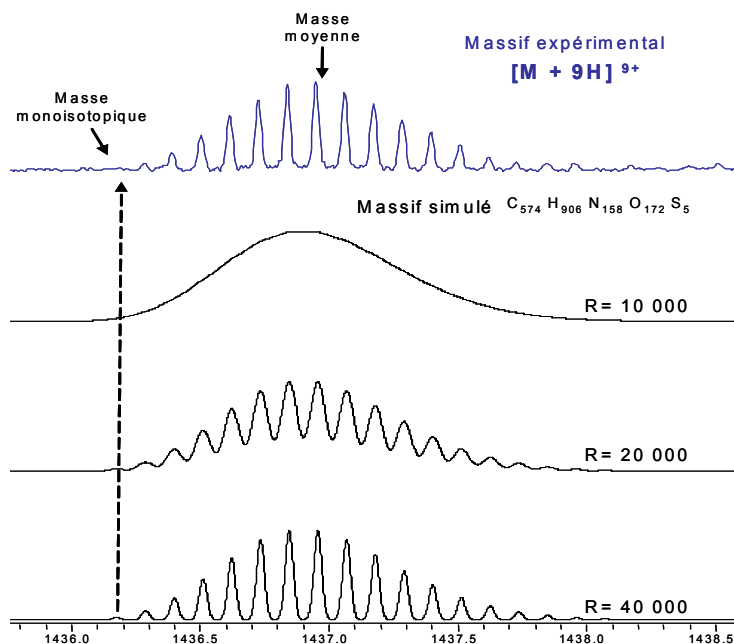


Figure 9: Détermination de la masse monoisotopique d'un composé protoné 9 fois à m/z 1436,072 – Comparaison avec des massifs simulés à différentes résolutions.

Bien que le nombre de masses détectées puisse sembler important, il est tout à fait clair que la liste des différentes masses que nous pu établir est loin d'être exhaustive et qu'en réalité le venin contient certainement bien plus que quatre-vingts composés. Outre le problème de gamme dynamique lié à l'analyseur, il est connu que l'utilisation de l'électrospray induit des phénomènes de suppression ionique. Ceci est amplifié par la large gamme de masse des composés présents dans le venin mais aussi par la diversité de leurs propriétés physico-chimiques. L'utilisation du nanoESI, qui permet une meilleure efficacité d'ionisation, limite ce problème. Néanmoins, la suppression ionique représenterait un réel problème si notre but final était de caractériser *toutes* les espèces présentes dans le mélange, et ce, quelles que soient leurs propriétés (acides ou basiques, hydrophiles ou hydrophobes, peptides ou petits composés organiques). Or, notre objectif n'est pas de déterminer toutes les masses présentes dans l'échantillon mais plutôt d'obtenir une empreinte du venin et de rechercher, à l'aide de la spectrométrie de masse, une famille précise de toxines. Pour cette raison, nous pouvons considérer que la suppression ionique n'est pas ici d'une importance cruciale.

En comparant ces premiers résultats à l'étude menée sur *Atractaspis microlepidota microlepidota*³, il apparaît que la diversité de la gamme de masse est similaire. En analysant plus en détails la cartographie massique que nous avons établie, on peut observer qu'environ la moitié de toutes les masses relevées est comprise entre 2000 et 3500 Da. Cette gamme de

masse correspond à des peptides composés de 20 à 30 acides aminés, longueur caractéristiques des sarafotoxines recherchées. Aucune de ces masses ne correspond néanmoins à la masse d'une isoforme de sarafotoxine déjà décrite. Ceci s'explique par le fait que chaque type de serpent possède un venin différent et que, si les familles de composés sont souvent conservées au sein d'une même espèce, il est rare de retrouver deux toxines totalement identiques dans le venin de deux serpents différents.

Une autre remarque peut être faite sur ces masses présentes autour de 3000 Da. Elles correspondent aux ions les plus intenses dans le spectre de masse. Or, nous savons que les sarafotoxines sont connues pour être surexprimées dans ce type de venin, ce qui va en faveur de la présence de sarafotoxines parmi ces espèces. Afin d'obtenir de plus amples informations structurales sur ces composés, des expériences de réduction des ponts disulfures ont été menées sur le venin brut.

B. Analyse du venin brut après réduction des ponts disulfures

La Figure 10 présente le spectre obtenu après réduction des ponts disulfures. De nombreux composés sont de nouveau observés. Chaque pont disulfure réduit conduisant à une augmentation de 2 Da (transformation d'un groupement S-S en deux groupements S-H), la comparaison des masses relevées avant et après réduction permet, pour chaque espèce, de déterminer précisément le nombre de ponts présents. A titre d'exemple, le décalage obtenu pour les composés triplement chargés majoritaires (vers m/z 1000) est également présenté *Figure 11*.

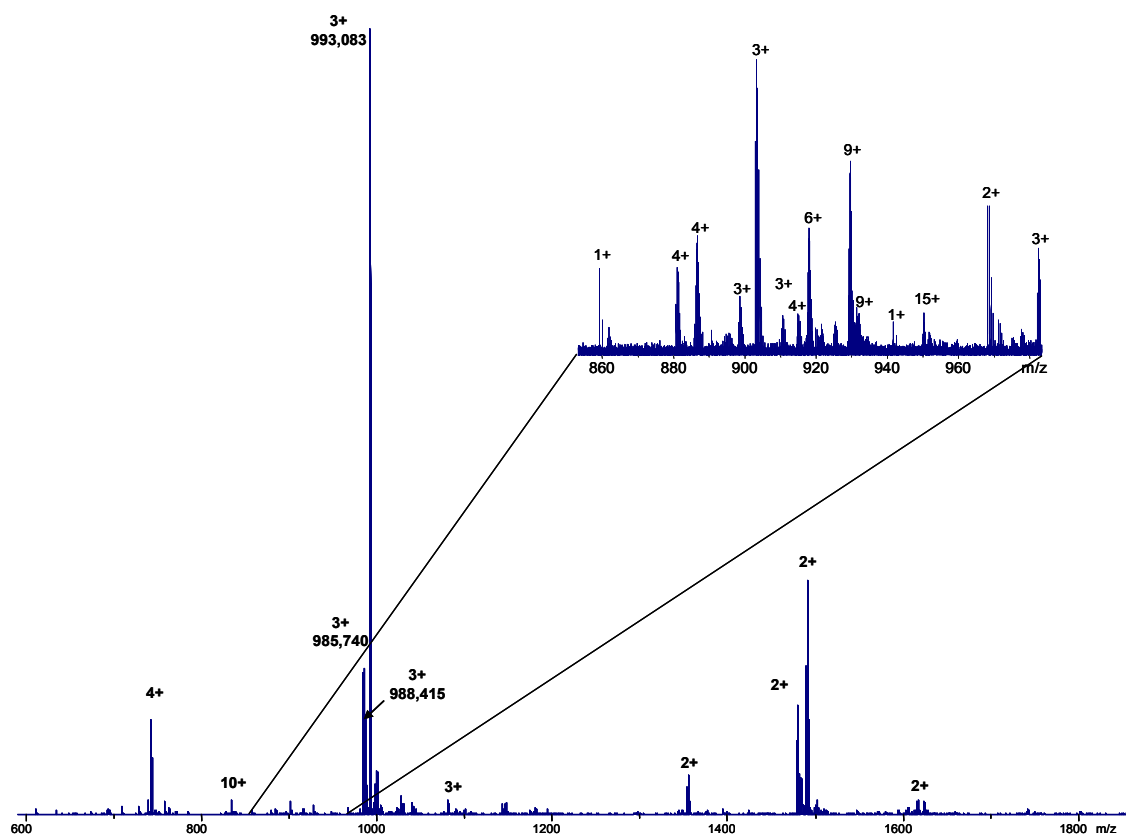


Figure 10 : : Spectre nanoESI-FTICR obtenu pour le venin brut après réduction des ponts disulfures

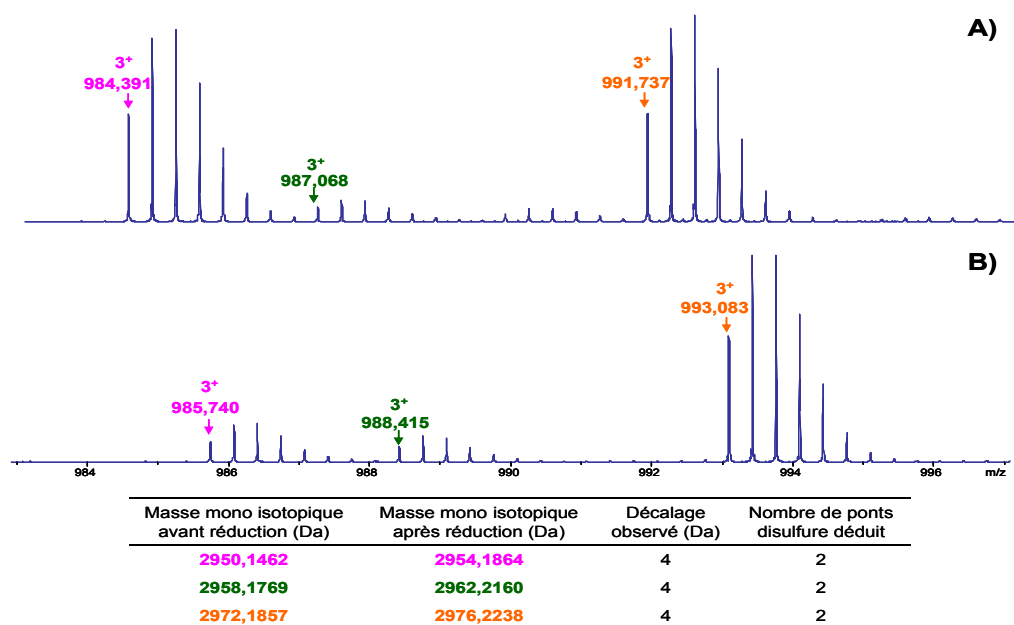


Figure 11: Réduction des ponts disulfures des composés majoritaires observés lors de l'analyse du venin brut– (A) avant réduction (B) après réduction des ponts disulfures au TCEP

De nombreux autres décalages en masse ont été détectés et répertoriés dans le *Tableau 2*. La plupart des masses est déplacée de 4 Da (34 masses sur 55), suggérant la présence de deux

ponts disulfures au sein de leur séquence initiale, comme dans celles des sarafotoxines. Il est remarquable que ces différences de masses affectent principalement la gamme de masse 1,9-2,5 kDa.

Masses moléculaires avant réduction (Da)	Masses moléculaires après réduction (Da)	Décalage observé en Da	Nombre de ponts S-S	Masses moléculaires avant réduction (Da)	Masses moléculaires après réduction (Da)	Décalage observé en Da	Nombre de ponts S-S
614,3035	614,3068	0,0	0	2994,2129	2998,2157	4,0	2
636,2991	636,3024	0,0	0	3002,1814	3006,2176	4,0	2
655,3374	655,3416	0,0	0	3004,1853	3008,2360	4,1	2
730,3340	730,3373	0,0	0	3010,1361	3014,1771	4,0	2
858,3933	858,3971	0,0	0	3018,2025	3020,1896	2,0	1
1419,6683	1421,6844	2,0	1	3085,2778	3089,3119	4,0	2
1833,5666	1837,6033	4,0	2	3293,3572	3297,3991	4,0	2
1929,8034	1933,8411	4,0	2	3353,4441	3357,4967	4,1	2
1935,6807	1939,7204	4,0	2	3515,4929	3519,5338	4,0	2
2027,7254	2031,7667	4,0	2	3537,5296	3541,5829	4,1	2
2050,7077	2054,7464	4,0	2	4606,8665	4606,8669	0,0	0
2076,6872	2080,7220	4,0	2	4633,8773	4633,8831	0,0	0
2175,7575	2179,7866	4,0	2	4703,9061	4703,9155	0,0	0
2280,8160	2284,8523	4,0	2	4730,9292	4730,9201	0,0	0
2288,8368	2292,8768	4,0	2	4774,9651	4774,9609	0,0	0
2296,8174	2300,8322	4,0	2	5392,3077	5394,2158	1,9	1
2474,9249	2478,9490	4,0	2	5421,1969	5421,2020	0,0	0
2680,0071	2684,0451	4,0	2	5490,5389	5498,5137	8,0	4
2688,0291	2692,0670	4,0	2	6976,9661	6985,0444	8,1	4
2702,0493	2706,0791	4,0	2	7612,3364	7620,4120	8,1	4
2926,2049	2928,2183	2,0	1	8343,9918	8352,0782	8,1	4
2940,2232	2942,2370	2,0	1	12917,4319	12921,5302	4,1	2
2950,1462	2954,1864	4,0	2	12950,4199	12954,4713	4,1	2
2958,1769	2962,2160	4,0	2	12981,9043	12981,4783	0,4	0
2966,1492	2970,1921	4,0	2	12983,2015	12987,2137	4,0	2
2972,1857	2976,2238	4,0	2	13274,3187	13280,4375	6,1	3
2982,1409	2984,1850	2,0	1	14215,2733	14231,5071	16,2	8
2988,1902	2992,1345	3,9	2				

Tableau 2: Liste des masses avant et après réduction du venin d'*A. irregularis* et nombre de ponts disulfures associés – les toxines possédant deux ponts S-S (sarafotoxines ?) sont encadrées

La comparaison des spectres obtenus avant/après réduction semble indiquer que plusieurs ions fortement chargés avant réduction apparaissent, après la réduction, avec des états de charge plus faibles, même si cela n'est pas le cas des molécules les plus intenses du spectre (forme $[M+3H]^{3+}$ majoritaire avant et après réduction). Nous avons déjà pu faire cette observation lors d'études d'autres venins ou de toxines purifiées. Ce phénomène va à l'encontre de ce qui est observé pour les protéines, puisqu'il est connu que dans la plupart des cas, le fait de réduire les ponts disulfures d'une protéine conduit à l'obtention de structures plus « dépliées » dans lesquelles de nouveaux sites susceptibles d'être protonés sont devenus accessibles. Il n'est donc pas rare d'observer des états de charge plus importants pour les protéines réduites en comparaison avec leur forme oxydée. Un phénomène similaire est observé lorsque l'on compare les états de charge observés pour une protéine analysée en milieu non dénaturant (protéine repliée, nombreux sites non accessibles pour la protonation) à ceux de la même protéine en conditions dénaturantes (protéine dépliée, nombreux sites accessibles). Dans le premier cas, la distribution d'état de charges observée est étroite et composée d'espèces peu chargées alors que dans le second cas, la distribution est beaucoup plus étalée et composée de nombreuses espèces fortement chargées.

Néanmoins, il est très difficile de comparer le comportement de protéines et celui de peptides. En effet, il n'est pas évident que la réduction des ponts S-S dans le cas d'espèces de petite taille (les sarafotoxines sont des peptides de 2,5-3 kDa environ) conduise à une accessibilité meilleure de certains sites de protonation. Pour expliquer nos observations, il paraît alors plus judicieux de comparer les affinités protoniques (AP) des différents groupements mis en jeu dans la réduction. Une AP plus élevée des liaisons disulfures par rapport aux thiols libres serait une explication plausible. Ne disposant pas de données thermodynamiques pour nos systèmes, on peut comparer les AP du dimethyldisulfure (qui modélise le pont S-S) et du methanethiol (qui représente le groupement -SH). Le dimethyldisulfure a une AP de $815,3 \text{ kJ.mol}^{-1}$ alors que celle du methanethiol n'est que de $773,4 \text{ kJ.mol}^{-1}$, ce qui indique (hors phénomènes de solvation) qu'un peptide possédant un pont S-S serait un meilleur « capteur » de protons que le même peptide dont le pont a été réduit. Ceci peut donc expliquer nos résultats.

Ces diverses expériences de spectrométrie de masse nous ont permis, d'une part de réaliser une cartographie massique du venin et d'autre part de cibler des toxines qui par leur masse et leur nombre de ponts disulfures possèdent un profil typique de sarafotoxines. Cependant, nous n'avons, à ce stade, aucune indication sur leur séquence et c'est pourquoi, nous avons complété ce travail par deux types d'expériences différentes : le clonage moléculaire des précurseurs des sarafotoxines afin de déduire les séquences en acides aminés de sarafotoxines potentielles et le séquençage *de novo* des espèces réduites, étant potentiellement des sarafotoxines par spectrométrie de masse FT-ICR.

IV. Recherche de sarafotoxines dans le venin

A. Approche par clonage moléculaire

A.1. Clonage de deux nouveaux précurseurs de sarafotoxines

La Figure 12 montre la structure de deux molécules d'ARN messager qui ont été clonées. Pour assurer le clonage de la séquence qui code pour les sarafotoxines matures, les amorces utilisées représentent des régions localisées avant le peptide signal et après le codon stop.

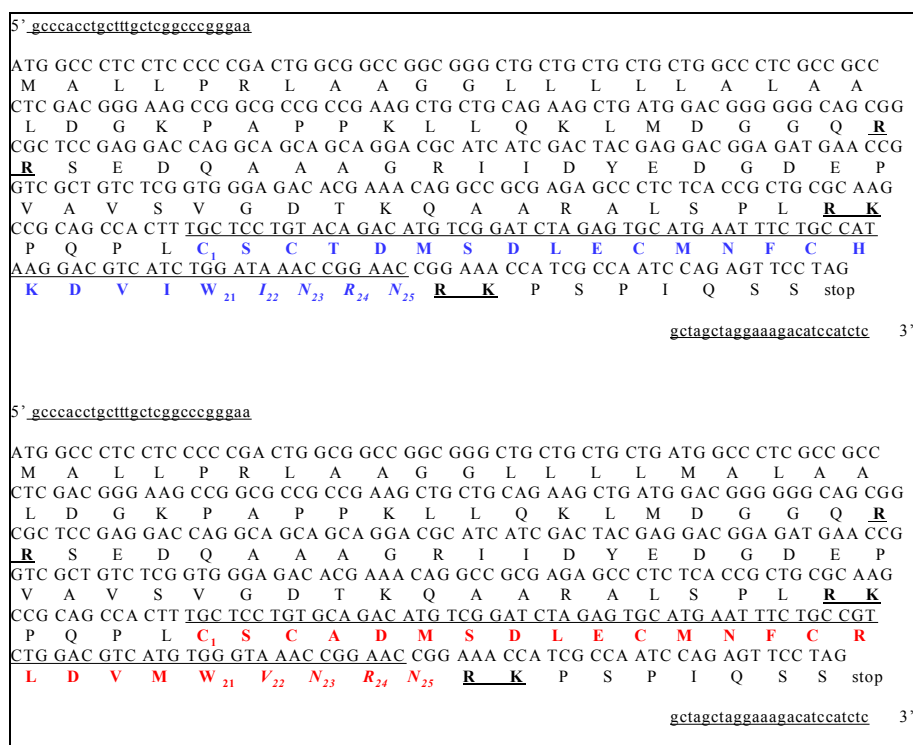


Figure 12: Séquences nucléotidiques et déduites en acides aminés de deux précurseurs des sarafotoxines d'*A. irregularis*

Entre les deux amorces, une séquence nucléotidique de 409 paires de bases correspondant à un précurseur protéique de 118 acides aminés apparaît clairement. En position C-terminale, on note la présence d'une séquence de sarafotoxine grâce au triplet CSC. Ce résultat indique sans ambiguïté que le venin d'*A. irregularis* contient des sarafotoxines. La séquence totale en acides aminés montre que le précurseur protéique commence par une méthionine d'initiation suivie d'un peptide hydrophobe. Ce peptide est tout à fait caractéristique d'une séquence signal, c'est-à-dire d'un peptide qui va permettre à la protéine (ou peptide) synthétisée d'être sécrétée. Le site de clivage potentiel de ce peptide signal a pu être localisé juste après le résidu glycine en position +23 du précurseur protéique¹⁹.

Plusieurs observations peuvent être apportées à la lecture de ce précurseur. Tout d'abord, entre la méthionine d'initiation et le début de la séquence SRTX, nous notons la présence d'un pro-peptide de 61 acides aminés, bien plus long que ceux notés lors des études précédentes. De plus, le codon stop n'intervient que 13 aminoacides après le tryptophane invariant des SRTX/Endothélines de la position +21, indiquant la possibilité que des sarafotoxines longues puissent exister. Enfin, trois doublets dibasiques sont localisés au sein du précurseur, doublets qui pourraient constituer des sites de clivages potentiels pour des

enzymes de type trypsine. Ces enzymes sont connues pour intervenir dans les procédés de maturation de nombreux précurseurs comme lors des maturations des pro-enképhalines en enképhalines²⁰ ou bien des pro-endothélines en big-endothélines^{11,21}. Le doublet situé après l'asparagine notée 25, est particulièrement intéressant puisqu'il indique un site de clivage potentiel juste après la séquence de la sarafotoxine, suggérant la présence de **sarafotoxines de 25 acides aminés**, qui représenteraient alors les plus longues sarafotoxines jamais décrites.

Le résultat le plus surprenant est que ce précurseur de sarafotoxines ne possède pas la même organisation que les autres précurseurs clonés jusqu'à présent à partir d'*A. engaddensis*⁹ et *A. microlepidota microlepidota*³. En effet, alors que pour ces deux serpents, les précurseurs s'étaient révélés polycistroniques, il apparaît nettement que dans le cas d'*A. irregularis*, seule une copie de sarafotoxine mature émerge de chaque précurseur. Ceci constitue la première description de deux organisations différentes pour des précurseurs codant des isoformes appartenant à la même famille de peptides.

Les très grandes similitudes des séquences en acides aminés pour les deux précurseurs clonés est présentée sur la *Figure 13*. Sur les 118 acides aminés qui les composent, seulement 6 sont différents, ce qui correspond à une homologie de séquence de 94,9%. Plus précisément, sur ces six acides aminés variants, cinq se situent dans les zones correspondant à une séquence de sarafotoxine.

```

MALLPRLAAG GLLLLLALAA LDGKPAPPKL LQKLMGGQR RSEDQAAAGR IIDYEDGDEP ...
_____M_____
... VAVSVGDTKQ AARALSPLRK PQPLCSC TDM SDLECMNFCH KDVIWINRNR KPSP IQSS stop
... _____A_____R L M V_____ stop

```

Figure 13: Comparaison des séquences en acides aminés des deux précurseurs clonés d'*A. irregularis*

A.2. Comparaison des résultats du clonage et de nanoESI-FTICR

A partir des deux séquences de sarafotoxines déduites des deux précurseurs, les masses exactes de chaque peptide commençant par le tag CSC ont été calculées. Les masses obtenues ont ensuite été comparées aux masses de la cartographie effectuée par nanoESI-FTICR. Il apparaît clairement que de nombreuses masses déterminées à partir du venin brut correspondent à celles calculées à partir des peptides commençant par CSC. Les différences

entre les masses théoriques et les masses expérimentales ont été calculées et se situent toutes entre 0,5 et 4,7 ppm, ce qui représente une première validation des résultats (*Tableau 3*).

Masses monoisotopiques expérimentales(Da)	Séquence des nouvelles sarafotoxines	Nombre d'acides aminés	Masses théoriques (Da)	ΔMr ppm	Identification
1833,5666	CSCTD MSDLE CMNFC	15	1833,5595	3,8	SRTX-i1 ₁₅
2076,6872	CSCTD MSDLE CMNFC HKD	18	2076,6814	2,7	SRTX-i1 ₁₈
2175,7575	CSCTD MSDLE CMNFC HKDVI	19	2175,7498	3,5	SRTX-i1 ₁₉
2288,8368	CSCTD MSDLE CMNFC HKDVI	20	2288,8339	1,3	SRTX-i1 ₂₀
2474,9249	CSCTD MSDLE CMNFC HKDVI W	21	2474,9132	4,7	SRTX-i1 ₂₁
2702,0493	CSCTD MSDLE CMNFC HKDVI WIN	23	2702,0402	3,4	SRTX-i1 ₂₃
2972,1857	CSCTD MSDLE CMNFC HKDVI WINRN	25	2972,1842	0,5	SRTX-i1₂₅
2988,1902	CSCTD MSDLE CMNFC HKDVI WINRN* a	25	2988,1791	3,7	SRTX-i1 ₂₅ *
3004,1853	CSCTD MSDLE CMNFC HKDVI WINRN** b	25	3004,1741	3,7	SRTX-i1 ₂₅ **
3353,4441	CSCTD MSDLE CMNFC HKDVI WINRN RKP	28	3353,4331	3,3	SRTX-i1 ₂₈
3537,5179	CSCTD MSDLE CMNFC HKDVI WINRN RKPSP	30	3537,5179	3,3	SRTX-i1 ₃₀
1935,6807	CSCAD MSDLE CMNFC RL	17	1935,6752	2,8	SRTX-i2 ₁₇
2050,7077	CSCAD MSDLE CMNFC RLD	18	2050,7021	2,7	SRTX-i2 ₁₈
2280,8160	CSCAD MSDLE CMNFC RLDVM	20	2280,8110	2,2	SRTX-i2 ₂₀
2296,8174	CSCAD MSDLE CMNFC RLDVM*	20	2296,8060	4,9	SRTX-i2 ₂₀ *
2680,0071	CSCAD MSDLE CMNFC RLDVM WVN	23	2680,0017	2,0	SRTX-i2 ₂₃
2950,1462	CSCAD MSDLE CMNFC RLDVM WVN RN	25	2950,1457	1,7	SRTX-i2₂₅
3515,4929	CSCAD MSDLE CMNFC RLDVM WVN RN RKPSP	30	3515,4794	3,8	SRTX-i2 ₃₀
2688,0291	CSCTD MSDLE CMNFC HKDVI WVN	23	2688,0245	1,7	SRTX-i3 ₂₃
2958,1769	CSCTD MSDLE CMNFC HKDVI WVN RN	25	2958,1686	2,8	SRTX-i3₂₅
2966,1492	CSCTD MSDLE CMNFC HKDVI WVN RN*	25	2966,1407	2,9	SRTX-i3 ₂₅ *
2982,1409	CSCTD MSDLE CMNFC HKDVI WVN RN**	25	2982,1356	1,8	SRTX-i3 ₂₅ **

Tableau 3 : Séquences primaires des deux familles de sarafotoxines déduites des clones SRTX - Les * indiquent les sarafotoxines retrouvées sous une forme oxydée (+16 Da), ** indique une double oxydation (+32 Da)

Les caractéristiques des différentes formes répertoriées dans le *Tableau 3* permettent de les classer en deux familles distinctes que l'on appellera SRTX-i1 et SRTX-i2, selon que les toxines proviennent du premier ou du second clone. Pour plus de lisibilité, le nombre d'acides aminés qui compose l'isoforme sera ajouté à la suite de son nom. A titre d'exemple, la toxine de 25 acides aminés appartenant à la famille SRTX-i1 sera nommée SRTX-i1₂₅. Le fait que le précurseur n'apparaisse pas exprimé sous la forme d'une seule et unique molécule mais sous celle d'un ensemble de molécules de différentes longueurs (de 15 acides aminés pour les plus courtes à 30 acides aminés pour les plus longues) est remarquable.

L'ionisation par electrospray étant une ionisation douce, il est peu vraisemblable que ces composés de longueurs variables résultent d'une fragmentation en phase gazeuse. On peut donc supposer que la présence de ces différentes formes reflète la diversité moléculaire naturelle du venin en sarafotoxines. Cette hypothèse est confortée par le fait que le même phénomène a été observé pour le venin d'*A. microlepidota microlepidota*³. Une explication plausible pourrait être la dégradation en solution d'un peptide initial de 30 acides aminés ou bien la présence de différents stades de maturation du pro-peptide initial.

Comme cela avait déjà été observé pour les autres venins, les sarafotoxines trouvées ici représentent les signaux les plus intenses du spectre de masse. Ainsi, bien que le précurseur soit monocistronique, les sarafotoxines semblent malgré tout surexprimées dans ce nouveau venin.

Pour compléter ces résultats de clonage, des expériences de fragmentation ont alors été menées sur les ions triplement chargés les plus intenses dans le spectre, *i.e.* ceux correspondant à SRTX-i1₂₅ et SRTX-i2₂₅.

B. Séquençage dans le venin brut

B.1. Confirmation de séquence pour les sarafotoxines clonées

Les deux sarafotoxines réduites SRTX-i1₂₅ et SRTX-i2₂₅ conduisent dans nos conditions expérimentales à des ions triplement chargés très intenses à m/z 993,030 et m/z 985,740 respectivement. Bien que les masses moléculaires correspondantes soient un peu élevées pour une fragmentation par collision, les expériences de SORI-CAD nous ont permis d'obtenir de bonnes couvertures de séquence, dans chacun des cas.

Le spectre de fragmentation obtenu pour l'ion à m/z 993,030 est reporté sur la *Figure 14*. Les différents ions fragments relevés sont répertoriés dans le *Tableau 4*. Pour améliorer le rapport signal/bruit, plusieurs dizaines de spectres consécutifs ont été accumulés. Les ions précurseurs ont été excités pendant 400 μ s hors résonance (500 Hz de décalage) en présence d'une pression de xénon de l'ordre de 10^{-7} mbar (5 pulses consécutifs).

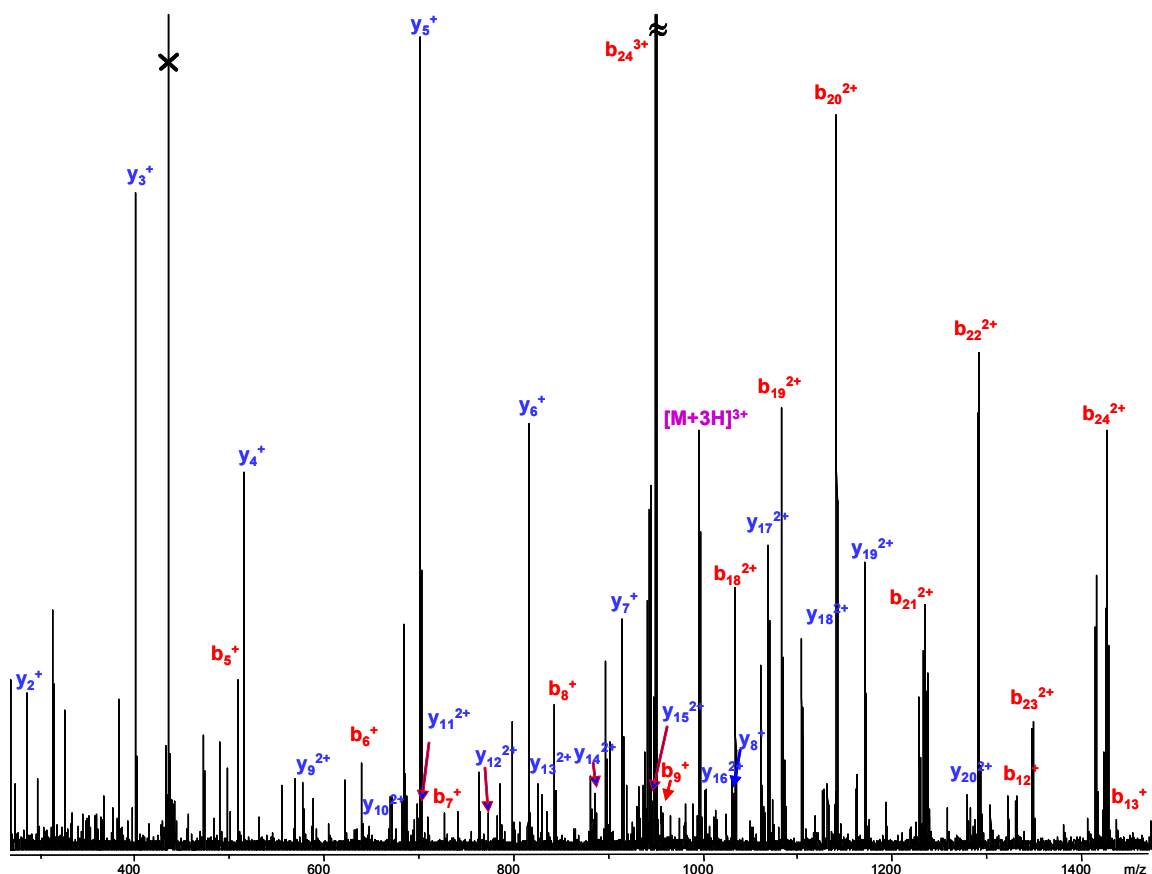
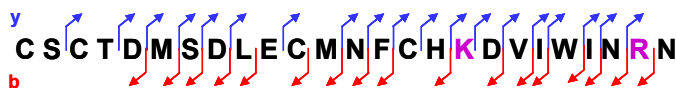


Figure 14 : Spectre MS/MS de l'ion à m/z 993,030 (3+) correspondant à SRTX M=2976,2238 Da (après réduction des ponts S-S au TCEP)

L'analyse du *Tableau 4* indique que le nombre de fragments obtenus est important et que la séquence de la sarafotoxine obtenue par les expériences de clonage moléculaire est confirmée. En effet, 21 liaisons peptidiques (sur 24) ont été clivées, ce qui correspond à une couverture de séquence de 87,5%. On peut remarquer que les fragments les plus intenses (monochargés) correspondent à des ions contenant l'extrémité C-terminale du peptide. La présence d'une arginine, résidu très basique pouvant séquestrer le proton, à cette extrémité explique parfaitement ce résultat.



Type d'ion	Masse expérimentale	Masse théorique	Erreur (ppm)	Type d'ion	Masse expérimentale	Masse théorique	Erreur (ppm)
y₂	289,1625	289,1613	4,15	b₅	510,1333	510,1323	1,96
y₃	403,2053	403,2043	2,48	b₆	641,1736	641,1728	1,25
y₄	516,2892	516,2883	1,74	b₇	728,2035	728,2048	1,79
y₅	702,3694	702,3676	2,56	b₈	843,2328	843,2317	1,30
y₆	815,4549	815,4517	3,92	b₉	956,3128	956,3158	3,14
y₇	914,5205	914,5201	0,44	b₁₁	1085,3601	1085,3584	1,57
y₈	1029,5531	1029,5471	5,83	b₁₂	1319,4134	1319,4081	4,02
y₉	1157,6451	1157,6420	2,68	b₁₃	1433,4606	1433,4510	6,70
y₁₀	1294,7004	1294,7009	0,39	b₁₄	1580,5246	1580,5194	3,29
y₁₁	1397,7099	1397,7101	0,14	b₁₆	1820,5777	1820,5875	5,38
y₁₂	1544,7764	1544,7785	1,36	b₁₈	2063,7173	2063,7094	3,83
y₁₃	1658,8206	1658,8215	0,54	b₁₉	2162,7718	2162,7778	2,77
y₁₅	1892,8745	1892,8711	1,80	b₂₀	2275,8533	2275,8619	3,78
y₁₇	2135,0049	2134,9978	3,33	b₂₁	2461,9457	2461,9412	1,83
y₁₈	2250,0343	2250,0247	4,27	b₂₂	2575,0302	2575,0253	1,90
y₁₉	2337,0625	2337,0568	2,44	b₂₃	2689,0783	2689,0682	3,76
y₂₀	2468,1064	2468,0972	3,73	b₂₄	2845,1709	2845,1693	0,56
y₂₁	2583,1143	2583,1242	3,83				
y₂₃	2787,1932	2787,1810	4,38				

Tableau 4: Assignment des ions **y** et **b** obtenus dans le spectre MS/MS de l'ion m/z 993,030 (3+)

La précision en masse obtenue sur les fragments est comprise entre 0,5 et 7 ppm, ce qui est satisfaisant pour des expériences MS/MS. La qualité du spectre obtenu (le nombre de fragmentations important) peut s'expliquer par trois raisons principales. La première est liée à l'état de charge fragmenté (3+) : la répulsion coulombienne due à la présence des trois protons sur le peptide permet d'obtenir de nombreuses fragmentations avec peu d'énergie supplémentaire. La seconde raison vient du fait que l'argon classiquement utilisé pour la fragmentation a été remplacé par du xénon, ce qui augmente l'énergie disponible au centre de masse. Enfin, il faut noter que les sites basiques sont bien répartis dans la séquence du peptide (amine N-terminale, lysine en position 17, arginine en position 24 et que l'excès de protons au départ évite au proton mobile responsable de la fragmentation de rester séquestré dans la séquence. L'abondance de l'ion b_{24} s'explique aisément par la localisation préférentielle du proton sur l'arginine. L'intensité importante de l'ion parent joue également un rôle favorable dans l'observation de fragments intenses.

Au final, SRTX-il₂₅, forme la plus abondante dans le spectre, possède quatre acides aminés supplémentaires par rapport aux SRTXs découvertes dans le venin d'*A. engaddensis* et un de plus que celles issues d'*A. microlepidota microlepidota*. Cependant, l'élongation observée dans notre cas est différente de celle observée chez *microlepidota* : W-I-N-R-N₂₅ vs D/E-Q/N-P/V₂₄. Ceci met en évidence une diversification supplémentaire de la composition en acides aminés des sarafotoxines. En effet, trois nouveaux acides aminés sont trouvés à des positions jusqu'alors très conservées : (i) une thréonine en position +4, (ii) une leucine en position +9 et (iii) une lysine en position +17.

Pour la seconde toxine, le spectre MS/MS de l'ion à m/z 985,740, correspondant à une masse moléculaire de 2954,1854 Da, confirme la présence d'une nouvelle sarafotoxine directement reliée au deuxième précurseur cloné puisque la partie C-terminale du peptide ainsi que quelques ions caractéristiques sont retrouvés (Tableau 5).



B.2. Découverte d'une nouvelle famille de sarafotoxines

Un troisième ion, à m/z 988,414, correspondant à un composé pouvant être une sarafotoxine potentielle, a alors attiré notre attention. La comparaison entre la masse moléculaire de ce composé, $M=2958,1769$ Da, et celle de SRTX-i1₂₅ ($M=2972,1857$ Da), fait apparaître une différence de masse de 14,0088 Da. Cet écart pourrait correspondre à la masse d'un groupement méthylène ($M_{CH_2} = 14,0157$ Da). Néanmoins, la précision en masse de la FT-ICR, permet d'émettre plusieurs autres hypothèses. En effet, en considérant une simple substitution d'un acide aminé par un autre, six possibilités existent : une leucine ou une isoleucine ($M=113,0841$ Da) par une valine ($M=99,0684$ Da), une alanine ($M=71,0371$ Da) par une glycine ($M=57,0215$ Da), une glutamine ($M=128,0586$ Da) par une asparagine ($M=114,0429$ Da), un acide glutamique ($M=129,0426$ Da) par un acide aspartique ($M=115,0269$) et une sérine ($M=87,0320$ Da) par une thréonine ($M=101,0477$ Da). Même si seulement 4 de ces possibilités sont envisageables pour la SRTX-i1 (qui ne possède ni glutamine, ni alanine), ces hypothèses ne prennent pas en compte les autres possibilités (modifications post-traductionnelles, plusieurs mutations...). Des expériences de fragmentation ont donc été entreprises sur l'ion triplement chargé à m/z 988,414 ($M=2962,2160$ Da). Le spectre de fragmentation permet de mettre en évidence la partie C-terminale du peptide inconnu et les ions obtenus sont répertoriés dans le *Tableau 6*.



Type d'ion	Masse expérimentale	Masse théorique	Erreur (ppm)
<i>y</i> ₂	289,1627	289,1613	4,84
<i>y</i> ₃	403,2057	403,2043	3,47
<i>y</i> ₄	502,2741	502,2727	2,79
<i>y</i> ₅	688,3533	688,352	1,89
<i>y</i> ₆	801,437	801,436	1,25
<i>y</i> ₇	900,5059	900,5045	1,55
<i>y</i> ₁₇	2120,9818	2120,9821	0,14
<i>y</i> ₁₉	2323,0242	2323,0411	7,27
<i>y</i> ₂₀	2454,1033	2454,0816	8,84
<i>b</i> ₄	395,1027	395,1054	6,83
<i>b</i> ₁₈	2063,6999	2063,7094	4,60
<i>b</i> ₁₉	2162,7915	2162,7778	6,33
<i>b</i> ₂₀	2275,8533	2275,8619	3,78
<i>b</i> ₂₁	2461,9275	2461,9412	5,56
<i>b</i> ₂₂	2561,0271	2561,0096	6,83
<i>b</i> ₂₄	2831,1396	2831,1537	4,98

Tableau 6: Assignement des ions *y* et *b* obtenus par SORI-CAD de l'ion à m/z 988,414 correspondant à une masse de 2962,2160 Da

L'analyse du spectre MS/MS obtenu indique clairement que la différence de 14,009 Da observée peut être corrélée à la substitution d'une isoleucine par une valine en position +22. Ceci apparaît nettement lorsque l'on calcule la différence entre les ions $b_{22}-b_{21}$ ainsi que y_4-y_3 , qui conduisent à des masses très voisines de celle de la valine (99,0684 et 99,0996 Da). Cette nouvelle sarafotoxine est également retrouvée sous plusieurs formes à 23 et 25 acides aminés (Tableau 7). Ces nouvelles sarafotoxines ne diffèrent de la famille des SRTX-i1 qu'au niveau des derniers acides aminés de la séquence, il est de ce fait impossible de différencier les formes courtes (<21 acides aminés) des deux familles.

Finalement, il apparaît que les familles SRTX-i1 et SRTX-i3 sont très voisines d'un point de vue structural. Elles ne diffèrent que par un seul acide aminé, une isoleucine (SRTX-i1) ou une valine (SRTX-i3) à la position +22, juste après W_{21} conservé. Un point important à noter est que la découverte des sarafotoxines SRTX-i3 ne repose que sur les expériences de spectrométrie de masse et ne provient pas de résultats du clonage moléculaire effectué sur les ARNm des cellules des glandes à venin.

Masses monoisotopiques expérimentales (Da)	Séquence des nouvelles sarafotoxines	Nombre d'acides aminés	Masses théoriques (Da)	ΔMr ppm	Identification
2688,0291	CSCTD MSDLE CMNFC HKDVI WVN	23	2688,0245	1,7	SRTX-i3 ₂₃
2958,1769	CSCTD MSDLE CMNFC HKDVI WVN	25	2958,1686	2,8	SRTX-i3₂₅
2966,1492	CSCTD MSDLE CMNFC HKDVI WVN [*]	25	2966,1407	2,9	SRTX-i3 ₂₅ [*]
2982,1409	CSCTD MSDLE CMNFC HKDVI WVN ^{**}	25	2982,1356	1,8	SRTX-i3 ₂₅ ^{**}

Tableau 7: Séquences primaires de la famille de sarafotoxines SRTX-i3 obtenue par séquençage *de novo* – Les * indiquent les sarafotoxines retrouvées sous forme oxydée (+16 Da), ** indique une double oxydation (+32 Da)

Pour expliquer l'absence de clone correspondant à cette séquence, plusieurs hypothèses peuvent être émises. Une des explications serait d'envisager un problème au niveau du clonage d'un précurseur éventuel dû à la faible abondance de ce dernier ou bien à une mauvaise corrélation avec les amorces choisies. Ces explications semblent toutefois ne pas pouvoir être retenues puisque l'abondance de la SRTX-i3 est tout à fait comparable à celle de la SRTX-i2 et que les amorces ont prouvé leur efficacité lors du clonage des précurseurs de SRTX-i1 et SRTX-i2. L'explication la plus plausible serait que les peptides de la famille SRTX-i3 dérivent d'un ARN messager modifié comme cela a été proposé pour les différents isoformes de l' α -bungarotoxine présents dans le venin du serpent *Bungarus multicinctus*²². Même si les mécanismes de modifications post-transcriptionnelles (« RNA editing ») sont mal connus dans les cellules de serpents, ils correspondent chez les mammifères à une

déamination de cytosine et d'adénine pour former des uraciles et des inosines. Ces désaminations sont effectuées par des enzymes comme les désaminases APROBEC-1 (cytosine → adénine) et DRADA (uraciles → inosines)²³.

A partir de ces informations, la différence observée entre les deux familles de SRTXs pourrait s'expliquer simplement par la modification d'une adénine (A) en guanine (G) de la première base du codon de l'isoleucine (AU+U,C ou A) pour donner un codon de valine (GU+U,C ou A). Seule une étude de l'ADN génomique d'*A. irregularis* pourrait confirmer cette hypothèse.

C. Etude de la conformation tridimensionnelle de SRTX-i1₂₅ par RMN

Toute cette étude a été réalisée au laboratoire de RMN de l'Institut Pasteur dirigé par Muriel Delepierre. Les sarafotoxines trouvées dans le venin d'*Atractaspis irregularis* sont les plus longues jamais décrites pour un venin d'*Atractaspididae*. SRTX-i1₂₅ a été synthétisée chimiquement et des expériences de résonance magnétique nucléaire ont été menées (Figure 15). Ces expériences ont un double objectif. Tout d'abord, elles sont destinées à vérifier que la conformation tridimensionnelle du noyau de ces nouvelles SRTXs (cystéine +1 → tryptophane +21) n'est pas modifiée par la présence de nouveaux acides aminés et reste donc très voisine de la conformation connue de la SRTX-b. Par ailleurs, ces expériences sont réalisées afin d'évaluer les changements conformationnels qui peuvent être induits par les quatre acides aminés supplémentaires en position C-terminale: formation d'une hélice α par exemple.

Les différents résultats montrent que la SRTX-i1₂₅ adopte effectivement le même type de conformation que celle qui avait été déterminée pour la SRTX-b, c'est-à-dire une structure étendue au niveau de l'extrémité N-terminale des acides aminés +1 à +4, puis un agencement coudé entre les +5 et +8 suivi d'une hélice α entre les acides aminés +9 et +15 stabilisée par les deux ponts disulfures.

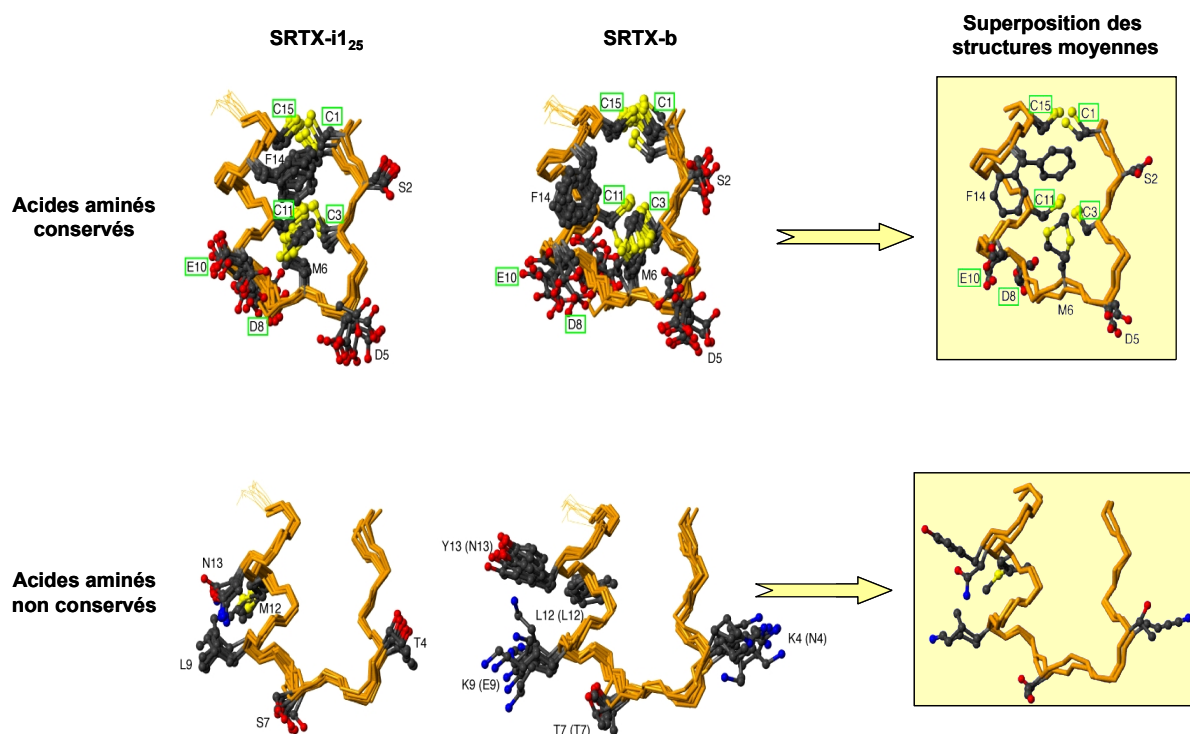


Figure 15: Étude de la conformation du noyau de la SRTX-i1₂₅ comparée à celle de la SRTX-b – Les figures en haut représentent les différentes structures superposées trouvées lors des analyses RMN pour chacune des deux toxines en se focalisant sur les acides aminés conservés– En bas, il s’agit de la même représentation mais en se focalisant cette fois sur les aminoacides non conservés - Les figures encadrées en jaune présentent la superposition des deux structures moyennes obtenues pour les deux toxines

Les deux représentations encadrées *Figure 15* illustrent bien ce point puisque les conformations des deux toxines sont quasiment superposables tant pour les acides aminés conservés que pour les autres. La conclusion est donc que l’ajout de ces acides aminés supplémentaires ne semble pas conduire à une structuration plus grande du système.

V. Conclusion et perspectives

Au cours de ce travail, nous avons montré que l’utilisation combinée de la spectrométrie de masse FT-ICR et du clonage moléculaire constitue un outil très efficace pour la caractérisation de nouvelles molécules bioactives directement dans une matrice aussi complexe qu’un venin brut, et ce, sans séparation chromatographique préalable. Les performances de la spectrométrie de masse FT-ICR en terme de précision sur la mesure de masse et haute résolution sont extrêmement utiles pour établir des cartographies massiques et comparer les masses mesurées à celles déduites du clonage ou encore pour le séquençage *de novo*.

Notre étude a révélé la présence de 22 nouvelles isoformes de sarafotoxines dont les chaînes en acides aminés varient de 15 à 30 acides aminés (Tableau 8). Ces toxines peuvent être regroupées en trois familles dont les espèces les plus représentées au sein du venin sont les formes composées de 25 acides aminés. Ces sarafotoxines sont plus longues que les endothélines (21 acides aminés) mais aussi que les autres sarafotoxines décrites lors des études sur *A. engaddensis*⁹ (21 acides aminés) et *A. microlepidota microlepidota*³ (24 acides aminés). Certains aminés qui apparaissaient conservés entre les endothélines et les sarafotoxines connues se sont révélés variables apportant une nouvelle diversification des séquences en acides aminés. Les tests biologiques sur ces nouvelles SRTXs n'ayant pas encore été réalisés, il est difficile d'évaluer l'influence de ces nouvelles modifications sur l'activité biologique.

		Alignements des séquences des endothélines avec celles de différentes sarafotoxines																										
ENDOTHELINES	Mammifères	ET-1	C	S	C	S	S	L	M	D	K	E	C	V	Y	F	C	H	L	D	I	I	W					
		ET-1 ₃₁	C	S	C	S	S	L	M	D	K	E	C	V	Y	F	C	H	L	D	I	I	W	V	N	T	P	E
		ET-2	C	S	C	S	S	W	M	D	K	E	C	V	Y	F	C	H	L	D	I	I	W					
		ET-2 ₃₁	C	S	C	S	S	W	M	D	K	E	C	V	Y	F	C	H	L	D	I	I	W	V	N	T	P	E
		ET-3	C	T	C	F	T	Y	K	D	K	E	C	V	Y	Y	C	H	L	D	I	I	W					
		ET-3 ₃₁	C	T	C	F	T	Y	K	D	K	E	C	V	Y	Y	C	H	L	D	I	I	W	I	N	T	P	E
	Rongeurs	VIC	C	S	C	N	S	W	L	D	K	E	C	V	Y	F	C	H	L	D	I	I	W					
	<i>A. engaddensis</i>	SRTX-a	C	S	C	K	D	M	T	D	K	E	C	L	N	F	C	H	Q	D	V	I	W					
		SRTX-b	C	S	C	K	D	M	T	D	K	E	C	L	N	F	C	H	Q	D	V	I	W					
		SRTX-c	C	T	C	N	D	M	T	D	E	C	L	N	F	C	H	Q	D	V	I	W						
SARAFOTOXINES	<i>A. bibroni</i>	Btx	C	S	C	A	D	M	T	D	K	E	C	L	Y	F	C	H	Q	D	V	I	W					
	<i>A. microlepidota microlepidota</i>	SRTX-m ₁₆	C	S	C	N	D	I	N	D	K	E	C	M	Y	F	C	H										
		SRTX-m ₂₄	C	S	C	N	D	I	N	D	K	E	C	M	Y	F	C	H	Q	D	V	I	W	D	E	P		
		SRTX-m ₁₂₄	C	S	C	N	D	M	N	D	K	E	C	M	Y	F	C	H	Q	D	V	I	W	D	E	P		
		SRTX-m ₂₁₁	C	S	C	N	D	I	N	D	K	E	C															
		SRTX-m ₂₁₈	C	S	C	N	D	I	N	D	K	E	C	M	Y	F	C	H	Q	D								
		SRTX-m ₂₂₄	C	S	C	N	D	I	N	D	K	E	C	M	Y	F	C	H	Q	D	I	I	W	D	E	P		
		SRTX-m ₃₂₄	C	S	C	N	D	M	N	D	K	E	C	V	Y	F	C	H	L	D	I	I	W	D	E	P		
		SRTX-m ₄₂₄	C	S	C	N	N	M	S	D	K	E	C	L	N	F	C	N	L	D	I	I	W	E	N	V		
	<i>A. irregularis</i>	SRTX-m ₅₂₄	C	S	C	N	D	M	N	D	K	E	C	V	Y	F	C	H	Q	D	I	I	W	D	E	P		
		SRTX-i ₁₅	C	S	C	T	D	M	S	D	L	E	C	M	N	F	C											
		SRTX-i ₁₁₈	C	S	C	T	D	M	S	D	L	E	C	M	N	F	C	H	K	D								
		SRTX-i ₁₁₉	C	S	C	T	D	M	S	D	L	E	C	M	N	F	C	H	K	D	V							
		SRTX-i ₁₂₀	C	S	C	T	D	M	S	D	L	E	C	M	N	F	C	H	K	D	V	I						
		SRTX-i ₁₂₁	C	S	C	T	D	M	S	D	L	E	C	M	N	F	C	H	K	D	V	I	W					
		SRTX-i ₁₂₃	C	S	C	T	D	M	S	D	L	E	C	M	N	F	C	H	K	D	V	I	W	I	N			
		SRTX-i ₁₂₅	C	S	C	T	D	M	S	D	L	E	C	M	N	F	C	H	K	D	V	I	W	I	N	R	N	
		SRTX-i ₁₂₈	C	S	C	T	D	M	S	D	L	E	C	M	N	F	C	H	K	D	V	I	W	I	N	R	N	R
		SRTX-i ₁₃₀	C	S	C	T	D	M	S	D	L	E	C	M	N	F	C	H	K	D	V	I	W	I	N	R	N	R
		SRTX-i ₂₁₇	C	S	C	A	D	M	S	D	L	E	C	M	N	F	C	R	L									
		SRTX-i ₂₁₈	C	S	C	A	D	M	S	D	L	E	C	M	N	F	C	R	L	D								
		SRTX-i ₂₂₀	C	S	C	A	D	M	S	D	L	E	C	M	N	F	C	R	L	D	V	M						
		SRTX-i ₂₂₃	C	S	C	A	D	M	S	D	L	E	C	M	N	F	C	R	L	D	V	M	W	V	N			
		SRTX-i ₂₂₅	C	S	C	A	D	M	S	D	L	E	C	M	N	F	C	R	L	D	V	M	W	V	N	R	N	
		SRTX-i ₂₃₀	C	S	C	A	D	M	S	D	L	E	C	M	N	F	C	R	L	D	V	M	W	V	N	R	N	R
		SRTX-i ₃₂₅	C	S	C	T	D	M	S	D	L	E	C	M	N	F	C	H	K	D	V	I	W	V	N			
		SRTX-i ₃₂₆	C	S	C	T	D	M	S	D	L	E	C	M	N	F	C	H	K	D	V	I	W	V	N	R	N	
	Acide aminés totalement conservés		C		C					D		E	C				C					W						
	Nouvelle diversité apportée						T				L							R	K		M				R	N	R	K

Tableau 8 : Alignement des séquences des sarafotoxines retrouvées dans le venin d'*A. irregularis* avec différentes séquences d'endothélines et de sarafotoxines d'autres *Atractaspidae*

Sept acides aminés sont conservés entre les trois nouvelles familles SRTX-i1, -i2 et -i3 : les quatre cystéines, qui interviennent de manière importante dans la structuration de tous les peptides, mais aussi des deux acides aspartiques (+8) et glutamique (+10) ainsi que le

tryptophane +21. On peut donc supposer que ces sept acides aminés, également conservés dans les séquences des endothélines, ont un rôle majeur pour l'activité des peptides vasoconstricteurs. Si les cystéines, impliquées dans deux ponts disulfures, assurent la structure compacte de ces toxines, les autres acides aminés pouvant jouer un rôle dans la liaison de ces toxines sur les récepteurs ET_A et ET_B. Ces hypothèses pourraient être vérifiées par des expériences de mutations successives des acides aminés conservés suivies de tests de liaisons aux récepteurs cibles.

Les expériences de RMN montrent que les variations observées sur ces nouvelles toxines (longueur de chaîne peptidique, nouveaux acides aminés...) ne modifient pas clairement leur structure tridimensionnelle qui reste très voisine de celles des premières sarafotoxines découvertes et des endothélines.

D'une manière inattendue, les précurseurs codant pour les sarafotoxines des familles SRTX-i1 et SRTX-i2 ont montré une organisation différente de celle qui avait été rencontrée lors des deux précédentes études. Ces résultats représentent la première observation, au sein de la même famille, de précurseurs présentant une organisation différente codant pour des toxines isoformes.

Les perspectives de ce travail sont de déterminer précisément l'activité biologique de ces nouvelles sarafotoxines et notamment de déterminer si les récepteurs ciblés sont les mêmes que ceux des autres sarafotoxines. Les premiers résultats obtenus suggèrent que les sarafotoxines longues (24 acides aminés) lient un site différent du récepteur des sarafotoxines à 21 acides aminés ou un nouveau récepteur. Cette question reste aujourd'hui ouverte. Malheureusement, peu d'équipes travaillent encore sur les récepteurs aux endothélines (notamment l'équipe israélienne qui a fait les premiers tests n'existe plus), ce qui ne facilite pas la suite de ce travail.

Références bibliographiques

- ¹ Kloog Y., Sokolovsky M., Similarities in mode and sites of action of sarafotoxins and endothelins, **1989**, *Trends Pharmacol. Sci.*, 10, 212-214.
- ² Bdolah A., Wolfberg Z., Kochva E., Snake Toxins-Sarafotoxins, a New Group of Cardiotoxic Peptides from the Venom of *Atractaspis*, **1991**, In: Harvey, A.L., (Ed), Pergamon Press, New York, p415.
- ³ Hayashi, M.A.F., Ligny-Lemaire C., Wollberg Z., Wery M., Galat A., Ogawa T., Muller B.H., Lamthanh H., Doljansky Y., Bdolah A., Stöcklin R., Ducancel F., Long-sarafotoxins: characterization of a new family of endothelin-like peptides, **2004**, *Peptides*, 25, 1243-1251
- ⁴ Deufel A., Cundall D., Feeding in *Atractaspis* (Serpentes : Atractaspididae) : a study in conflicting functional constraints, **2003**, *Zoology*, 106, 43-61.
- ⁵ Kochva E., Viljoen C.C., Botes D.P., A new type of toxin in the venom of the genus *Atractaspis* (Atractaspididae), **1982**, *Toxicon*, 20, 581-592.
- ⁶ Takasaki C., Tamiya N., Bdolah A., Wollberg Z., Kochva E., Sarafotoxins S6, several isotoxins from *Atractaspis engaddensis* (burrowing asp) venom that affect the heart, **1988**, *Toxicon*, 26, 543-548.
- ⁷ Kloog Y., Ambar I., Sokolovsky M., Kochva E., Wollberg Z., Bdolah A., Sarafotoxin, a novel vasoconstrictor peptide: phosphoinositide hydrolysis in rat heart and brain, **1988**, *Science*, 242, 268-270.
- ⁸ Becker A., Dowdle E.B., Hechler U., Kauser K., Donner P., Schleuning W.D., Bibrotoxin, a novel member of the endothelin/sarafotoxin peptide family from the venom of the burrowing asp *Atractaspis bibroni*, **1993**, *FEBS Lett.*, 315, 100-103.
- ⁹ Ducancel F., Matre V., Dupont C., Lajeunesse E., Wollberg Z., Bdolah A., Kochva E. Boulain J.C., Menez A., Cloning and sequence analysis of cDNAs encoding precursors of sarafotoxins, **1993**, *J. Biol. Chem.*, 268, 3052-3055.
- ¹⁰ Kochva E., Bdolah A., Wollberg Z., Sarafotoxins and endothelins: evolution, structure and function, **1993**, *Toxicon*, 31, 541-568.
- ¹¹ Yanagisawa M., Kurihara H., Kimura S., Tomobe Y., Kobayashi M., Yazaki Y., Goto K., Masaki T., A novel potent vasoconstrictor peptide produced by vascular endothelial cells, **1988**, *Nature*, 332, 411-415.
- ¹² Bloch K.D., Hong C.C., Eddy T.B., Shows T.B. and Quaternous T., cDNA cloning and chromosomal assignment of the endothelin 2 gene: vasoactive intestinal contractor in rat endothelin 2, 1991, *Genomics*, 10, 236-242.
- ¹³ Kishi, F., Minami.K., Okishima N., Murakami M., Mori S., Yano M., Niwa Y., Nakaya Y., Kido H., Novel 31-amino-acid-length endothelins cause constriction of vascular smooth muscle, **1997**, *Biochem. Biophys. Res. Com.*, 248, 387-390.
- ¹⁴ D'Orléans-Juste P., Plante M., Honoré J.C., Carrier E., Labonté J., Synthesis and degradation of endothelin-1, **2003**, *Can. J. Physiol. Pharmacol.*, 81, 503-510.
- ¹⁵ Waller D., Cudney R., Wolff M., Day J., Greenwood A., Larson S., Pherson A., Crystallization and preliminary X-ray analysis of human endothelin, **1992**, *Acta Cryst.*, 48, 239-240.
- ¹⁶ Andersen N. H., Chen C., Marschner T.M., Krystek S.R., Bassolino D.A., Conformational isomerism of endothelin in acidic aqueous media: a quantitative NOESY analysis, **1992**, *Biochem.*, 31, 1280-1295.

- ¹⁷ Tomoaki H., Kobayashi Y., Nishimura S., Ohkubo T., Kyogoku Y., Nakajima K., Solution conformation of endothelin determined by means of 1H-NMR spectroscopy and distance geometry calculations, **1991**, *Prot. Eng.*, 4, 509-518.
- ¹⁸ Atkins A.R., Martin R.C., Smith R., 1H-NMR studies of sarafotoxine SRTX-b, a non selective endothelin receptor agonist, and IRL 1620, an ETB receptor specific agonist, **1995**, *Biochem.*, 34, 2026-2033.
- ¹⁹ Van Heijne G., A new method for predicting signal sequence cleavage sites, **1986**, *Nucleic Acids Res.*, 14 (11), 4683-4690.
- ²⁰ Gubler U., Seeburg P., Hoffman B. J., Gage P.L., Udenfriend S., *Molecular cloning establishes proenkephalin as precursor of enkephalin-containing peptides*, **1982**, *Nature*, 295, 206-208.
- ²¹ Opgenorth T.J., Wu-Wong J.R., Shiosaki K., **1992**, Endothelin-converting enzymes, *FASEB J.*, 6, 2653-2659.
- ²² Liu L.F., Chang C.C., Liao, M.Y., Kuo, K.W., Genetic characterization of the mRNAs encoding a-bungarotoxin isoforms and RNA editing in Bungarus multicinctus gland cells, *Nucleic Acids Res.*, **1998**, 26, 5624-5629.
- ²³ Ashkenas J., Gene regulation by mRNA editing, *J. Am. Hum. Genet.*, **1997**, 60, 278-283.

Chapitre II

Séquençage *de novo* de conotoxines par
spectrométrie de masse - Apports du couplage
chromatographique

I. Introduction générale

En analyse protéomique, la fragmentation de peptides issus d'une digestion trypsique de la protéine à identifier permet d'obtenir des éléments de séquence primaires. L'utilisation de la trypsine permet d'obtenir des peptides d'une taille généralement raisonnable pour la dissociation (<2500 Da), dont le dernier acide aminé en position C-terminale, est soit une arginine, soit une lysine. La présence de ces sites basiques, capables de piéger un proton, oriente souvent la fragmentation et conduit à la formation de nombreux ions y dans les spectres MS/MS réalisés par collision basse énergie avec un gaz inerte. La présence de cette série d'ions y , et notamment d'un ion y_1 à m/z 147,1128 (Lys) ou m/z 175,1190 (Arg), permet une lecture facile de la séquence du peptide fragmenté en démarrant de l'acide aminé C-terminal.

Dans le cas de peptides naturels comme les conotoxines, qui ne portent pas de résidu basique sur leur extrémité C-terminale et qui sont fortement structurés et modifiés post-traductionnellement, le problème est bien plus délicat. Tout d'abord, la présence de nombreux ponts disulfures dans leur structure ne permet pas une fragmentation directe de ces composés. Une étape de réduction chimique de ces ponts est à intégrer dans la stratégie d'analyse d'autant que la détermination du nombre de ponts disulfures de ces toxines est importante. En effet, la *Figure 1* montre la relation étroite entre la masse des conotoxines, le nombre de ponts disulfures qu'elles contiennent et leur famille d'appartenance. Si on prend le cas des conotoxines possédant un unique pont disulfure, on observe qu'elles possèdent toutes une masse voisine de 1000 Da. Il s'agit d'une famille très intéressante composée des conopressines, actives sur les récepteurs à la vasopressine, et des contryphans dont les cibles biologiques sont encore mal définies. Jusqu'à aujourd'hui, peu de ses représentants sont connus et la découverte de nouveaux contryphans permettrait d'étoffer cette famille et d'ajouter de nouveaux outils nécessaires à la découverte de leur activité biologique. Un intérêt encore plus grand résiderait dans la découverte d'une toxine possédant un unique pont disulfure mais dans une gamme de masse totalement différente, par exemple vers 500 ou 2000 Da. Cela constituerait la signature unique d'une nouvelle famille de conotoxines. Dans la même optique, une nouvelle branche de la superfamille M a récemment été découverte¹ (notée sur le diagramme M-superfamily*). Cette famille possède une signature très intéressante du fait qu'il s'agit des plus petites conotoxines connues possédant trois ponts disulfures.

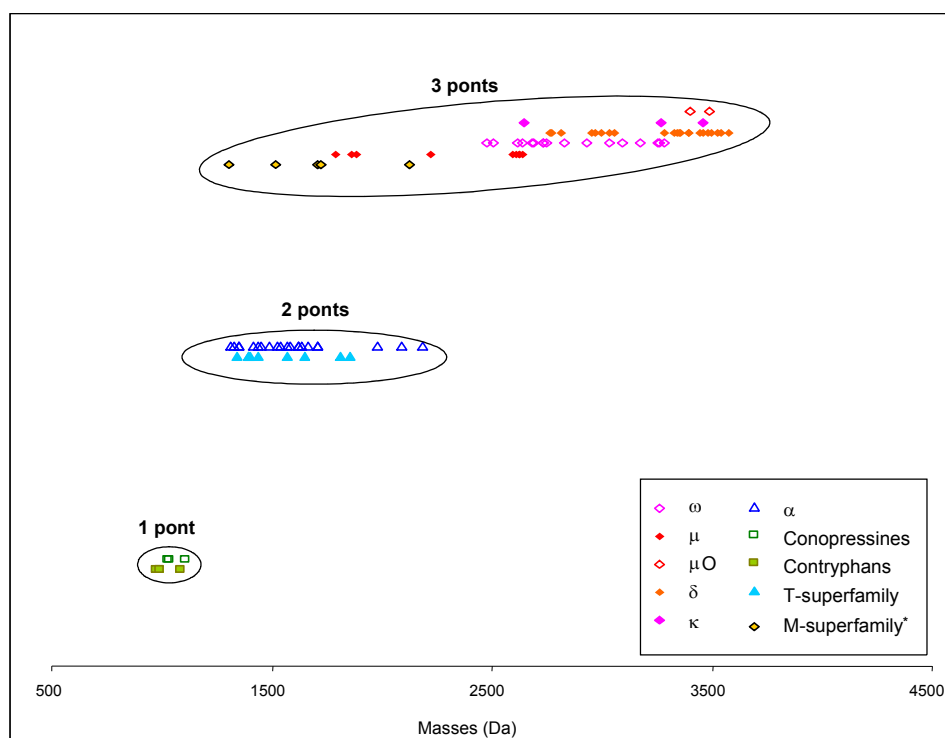


Figure 1: Répartition des masses de plusieurs familles de conotoxines en fonction du nombre de ponts disulfures dans leur séquence – seules les familles possédant 1, 2 ou 3 ponts sont répertoriées ici – la M-superfamily* indique une nouvelle branche de la M-superfamily composée jusqu’alors des μ -, Ψ - et κ M-conotoxines

L’absence d’arginine ou de lysine en C-terminal conduit à des spectres de fragmentation complexes, sans réel point de départ pour leur analyse du fait de l’absence des ions y_1 caractéristiques des peptides tryptiques. Il faut par ailleurs noter que la présence d’acides aminés modifiés (hydroxyproline par exemple) et de modifications post-traductionnelles (amidation de l’extrémité C-terminale, pyroglutamylation en N-terminal) ne facilite pas non plus ces études. Il convient également de rappeler que le génome des cônes marins n’est pas séquencé et que, par conséquent, obtenir un tag de séquence ne suffit pas pour identifier une conotoxine. Dans ce contexte, la précision en masse de la spectrométrie de masse FT-ICR est importante puisqu’elle permet : (i) dans le cas de spectres présentant des séries de pics discontinus, de proposer des *tags* de séquence basés uniquement sur des mesures de masse exacte des ions observés (ii) de trancher entre deux acides aminés de masse très proche (hydroxyproline/leucine ou isoleucine par exemple).

Une autre difficulté pour l’analyse de peptides naturels présents dans un venin vient du fait que les constituants d’un venin présentent une gamme dynamique très grande avec des

variabilités très importantes au niveau de la masse moléculaire des espèces. Ainsi, si certaines toxines majoritaires peuvent être étudiées directement par infusion d'une solution de venin brut, une séparation chromatographique est souvent nécessaire pour réaliser l'étude plus complète du protéome d'un venin. Les venins de cônes n'étant pas disponibles dans de grandes quantités, l'utilisation de techniques de chromatographie miniaturisées peut s'avérer nécessaire.

La première partie de ce chapitre est consacrée à l'étude du venin de *Conus virgo*. Nous montrons notamment qu'il est possible, par FT-ICR, de séquencer de petites conotoxines directement dans le mélange complexe et que les performances de la technique en terme de précision et de résolution sont utiles pour déterminer des séquences complètes. Dans une seconde partie, l'analyse du venin de *Conus ermineus* met en évidence que dans le cas de venins très complexes, il est indispensable d'avoir recours à une séparation chromatographique avant l'analyse par spectrométrie de masse, aussi résolutive soit-elle. Les avantages et limitations des différentes approches utilisées (infusion directe, couplage on-line et off-line, fragmentations dans un FT-ICR ou un Q-TOF) pour l'analyse d'un venin brut sont discutés.

II. Séquençage *de novo* de conotoxines dans le venin brut de *Conus virgo*

Conus virgo est un gastéropode marin que l'on trouve dans les mers chaudes (indopacifiques) et notamment dans les lagons de faible profondeur de Polynésie française et des Philippines. Son venin a jusqu'alors été très peu étudié mais une conotoxine, nommée ViTx², qui agit sur les canaux ioniques potassiques a été séquencée et étudiée récemment. La même équipe a menée une seconde étude sur ce cône et a notamment séquencé trois nouveaux précurseurs de conotoxines appartenant à la O-superfamily³.

A. Analyse du venin brut de *Conus virgo* par nanoESI-FT-ICR

A.1. Analyse du venin brut

Les premiers essais d'analyse en nanoESI sur le venin brut par simple dilution de la solution stock dans du méthanol ne permettent pas d'obtenir un spray, caractérisant la présence probable de sels dans l'échantillon. Une purification sur microcolonne de type ZipTip[®] a donc été introduite au niveau de la préparation d'échantillon permettant l'obtention d'un spray stable instantanément.

Le spectre du venin brut obtenu par infusion directe en nanoESI-FT-ICR (ACN/H₂O/AF, 49,95/49,95/0,1%) permet de détecter 64 composés dont les masses sont comprises entre 500 et 4500 Da (*Figure 2*).

Bien que les différentes espèces soient détectées sur une gamme de masse assez large, la majorité (environ 90%) d'entre elles sont inférieures à 4 kDa (*Figure 3*). Une analyse plus fine du spectre obtenu en nanoESI permet de visualiser que deux composés, ayant des masses moléculaires de 1328,5134 et 1358,5592 Da, sont largement prédominants. Ils sont retrouvés majoritairement sous des formes $[M+H]^+$ et $[M+2H]^{2+}$ accompagnées d'adduits $[M+Na]^+$ et $[M+K]^+$. La somme des intensités de leurs signaux correspondent respectivement à 18% et 51% du signal total (en se basant sur l'intensité du pic monoisotopique). Ces deux masses n'ayant jamais été trouvées dans aucun venin de cônes, nous avons décidé de caractériser les structures auxquelles elles correspondaient en réalisant la fragmentation directement à partir

du venin brut. Ceci était envisageable dans la mesure où les intensités dans le mélange complexe étaient relativement importantes.

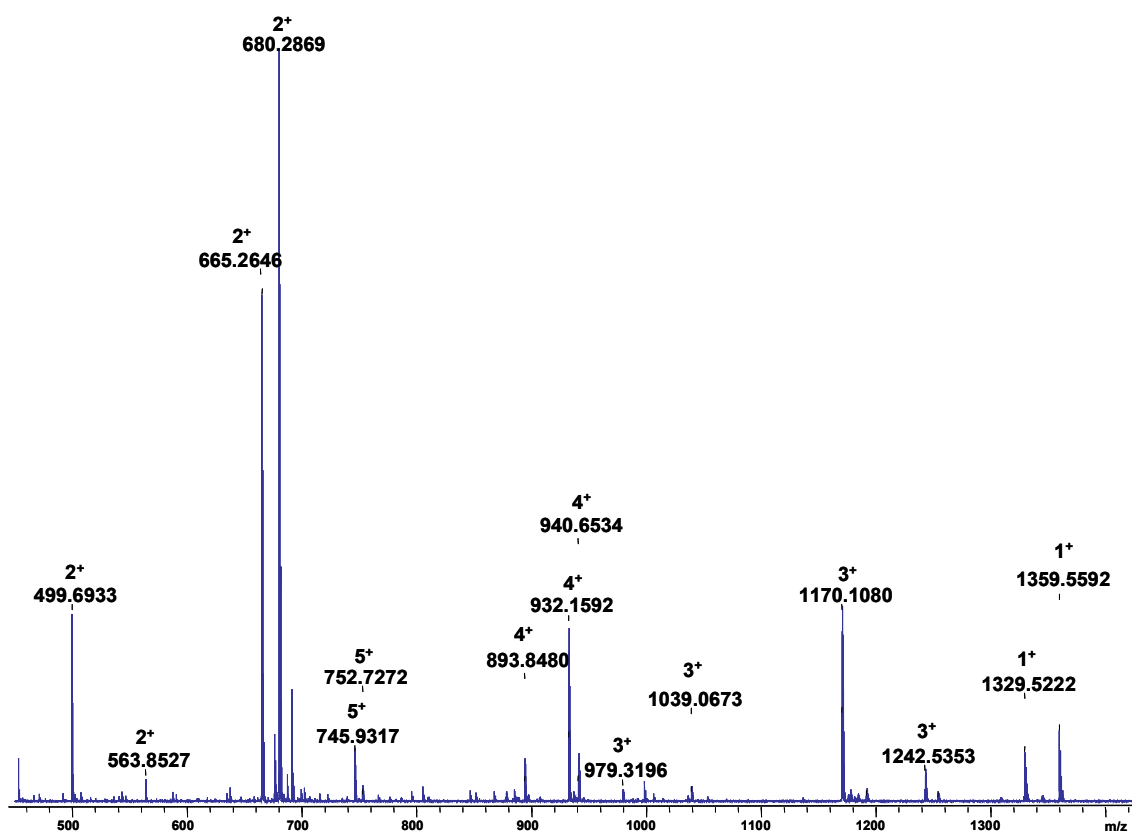


Figure 2: Spectre nanoESI-FT-ICR du venin brut de *Conus virgo*

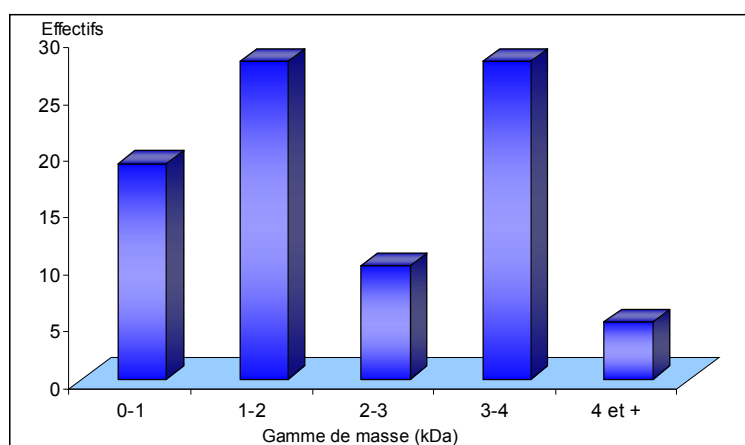


Figure 3: Répartition massique des composés présents dans le venin de *Conus virgo*

L'objectif de cette étude était double : déterminer la structure puis éventuellement évaluer l'activité biologique de ces nouvelles espèces (conotoxines ?) mais également évaluer l'intérêt

d'utiliser un spectromètre de masse FT-ICR pour ce type d'application, ce qui n'avait jamais été entrepris. La première étape a consisté à mesurer le plus précisément possible les masses de ces deux toxines potentielles. L'analyse du venin brut, en présence d'un mélange de peptides utilisés pour la calibration, a été effectuée par MALDI-FT-ICR. Le mélange de peptides utilisé et commercialisé par la société Bruker (*Pepmix*) est composé d'angiotensine I ($[M+H]^+$ 1296,6853), d'angiotensine II ($[M+H]^+$ 1046,5420), de substance P ($[M+H]^+$ 1347,7361) et de bombésine ($[M+H]^+$ 1619,8230). Le choix de ce mélange est judicieux puisque les masses des peptides de calibration encadrent parfaitement les masses de nos molécules inconnues. En utilisant ces peptides à une concentration permettant d'observer des signaux de même intensité que ceux de nos toxines, la précision sur la mesure de masse obtenue devrait être de l'ordre de la ppm.

Les masses moléculaires de nos toxines potentielles ainsi déterminées sont respectivement 1328,5142 et 1358,5592 Da, voisines de celles mesurées lors des expériences en nano-electrospray. La fragmentation directe des toxines ne peut être envisagée puisque, en tant que conotoxines potentielles, il existe de fortes chances qu'elles possèdent un ou plusieurs pont(s) disulfure(s). Des expériences de réductions de ces ponts ont donc été réalisées afin non seulement de rendre accessibles tous les acides aminés de la séquence peptidique lors des expériences MS/MS mais aussi afin de les dénombrer.

A.2. Analyse du venin réduit

Chaque pont disulfure réduit se traduit sur le spectre de masse par un gain de 2 Da consécutif à la réduction de la liaison S-S en deux liaisons S-H (*Figure 4*).

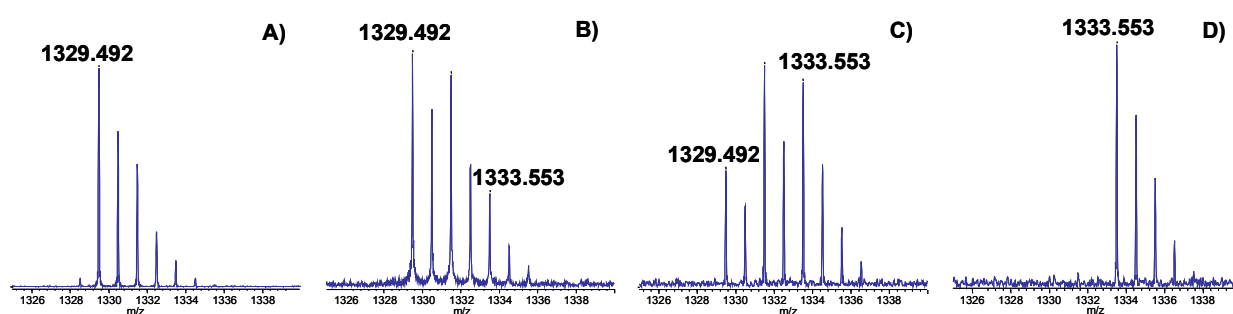


Figure 4: Réduction des ponts disulfures d'une toxine d'intérêt suivie par MALDI-FT-ICR - (A) à t=0 min (B) à t=10min, (C) à t=20 min et (D) à t=30 min

Les masses des deux toxines réduites sont mesurées à $[M+H]^+$ 1333,553 et $[M+H]^+$ 1363,598. Un gain de masse de 4 Da est donc observé pour chacune d'entre elles caractérisant la présence de deux ponts disulfures au sein de leur séquence et de ce fait, un minimum de 4 cystéines. Cette première observation qui pourrait paraître anodine est très informative et peut permettre de déterminer la famille d'appartenance de ces deux toxines. En effet, seules les α -conotoxines⁴ ainsi que la superfamille T⁵ (incluant les χ -conotoxines et les toxines possédant l'agencement des ponts de type V) sont aujourd'hui connues chez les cônes pour posséder des toxines présentant deux ponts disulfures. D'un point de vue biologique, les α -conotoxines ont une activité sur les récepteurs nicotiniques sensibles à l'acétylcholine (AChR)⁶. Les AChR sont impliqués dans de nombreux processus neurophysiologiques comme la mémoire et dans plusieurs maladies neuro-dégénératives comme les maladies d'Alzheimer et de Parkinson, la schizophrénie ou encore le syndrome de Gilles de la Tourette. Pour la superfamille T, les χ -conotoxines qui forment l'une de ses branches sont des inhibiteurs du transporteur de l'adrénaline⁷, un récepteur responsable de la terminaison de la transmission nerveuse. L'autre branche de cette famille, constituée de peptides présentant un arrangement des ponts S-S de type V n'a encore aucune activité bien définie⁵.

Quelle que soit la famille d'appartenance des deux toxines analysées dans le venin de *Conus virgo*, leur caractérisation présente un intérêt biologique. En effet, s'il s'agit d' α -conotoxines, il est envisageable de les utiliser comme outils pour l'étude des AChR avec, éventuellement, de nouvelles spécificités pour des sous-types de ces récepteurs. S'il s'agit d'une conotoxine T, en déterminer l'activité pourra apporter de nouvelles informations sur une classe de composés très mal connue.

B. Séquençage *de novo* de deux nouvelles conotoxines

Dans le but d'obtenir un maximum d'informations, les expériences de fragmentation ont été effectuées par SORI-CAD sur les espèces réduites. Alors que les espèces $[M+2H]^{2+}$ représentent les signaux les plus intenses du spectre nanoESI du venin brut, il apparaît qu'après l'étape de réduction, seules les espèces $[M+H]^+$ sont détectées, éliminant toute possibilité de fragmentation par capture d'électron dissociative.

B.1. Analyse des spectres MS/MS et séquençage *de novo*

B.1.a. Etude générale des spectres de fragmentation

Les expériences de MS/MS ont donc été effectuées par collision avec du xénon après sélection successive dans le spectre de chacune des deux toxines réduites (Figure 5).

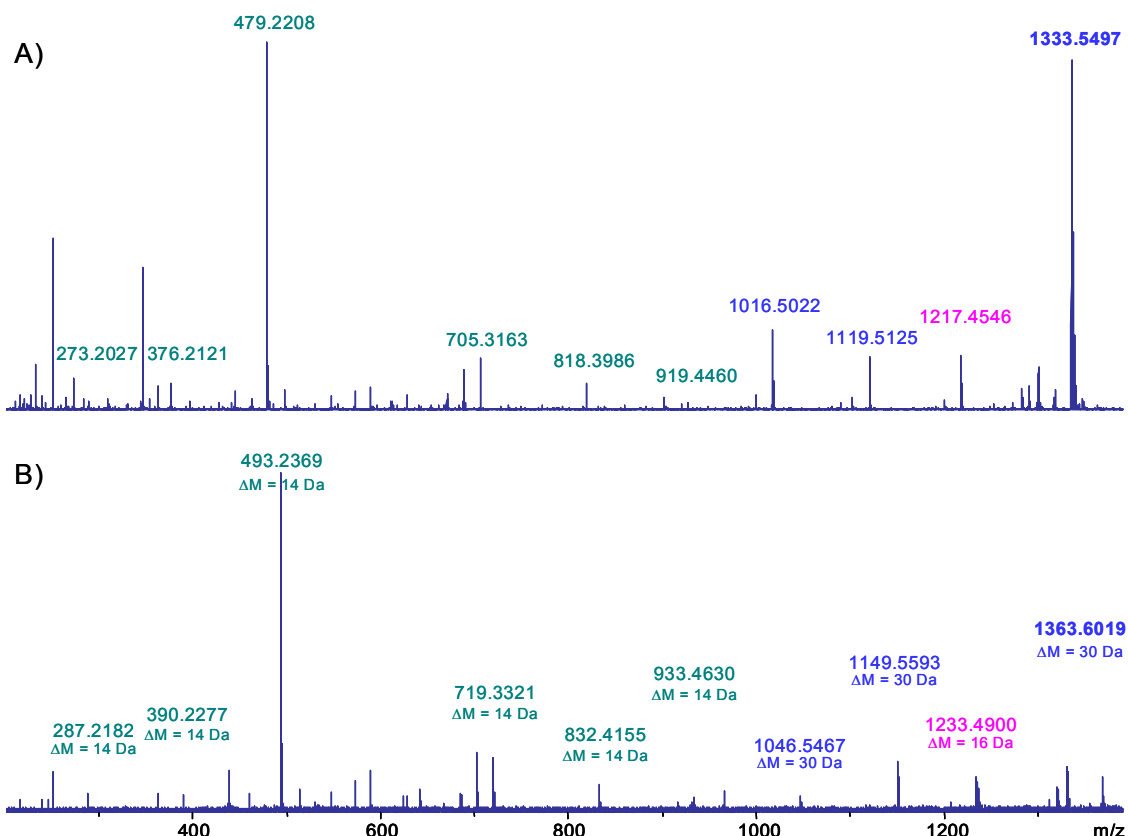


Figure 5: Les spectres MS/MS des deux toxines réduites de *Conus virgo* présentent de grandes similarités – (A) Toxine à m/z 1333,5497 et (B) m/z 1363,6019

La première observation est que les deux spectres obtenus sont très semblables. En les étudiant plus en détail, il apparaît que la différence de 30 Da observée entre les ions parents (m/z 1333,5497 et 1363,6038) est retrouvée entre différents ions fragments (en bleu), caractérisant une partie commune entre les deux toxines. De plus, il apparaît clairement que d'autres fragments se retrouvent séparés de 16 Da (en violet) ou de 14 Da (en vert), la somme de ces deux différences permettant de retrouver l'écart initial de 30 Da. Il est probable que ces deux différences traduisent le fait que ces deux toxines ont des séquences semblables et qu'elles se différencient uniquement par deux acides aminés (l'un apportant une différence de 14 Da et l'autre de 16 Da).

La recherche de morceaux de séquences (*tags*) a été effectuée automatiquement par le logiciel Biotoools de Bruker en imposant une incertitude sur les séquences trouvées de 0,02 Da (10-15 ppm). Cette recherche ne permet pas de retrouver toute la séquence. Néanmoins, deux tags sont proposés (*Figure 6*) à partir des ions les plus intenses du spectre.

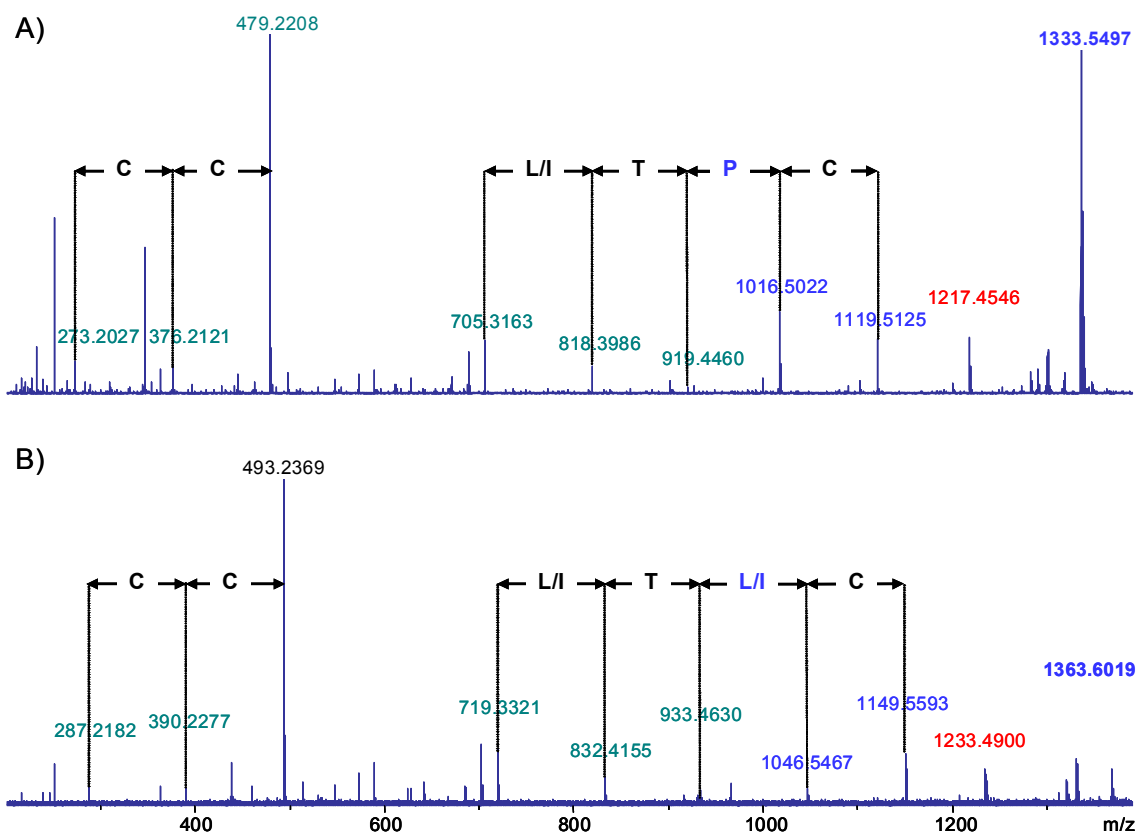


Figure 6: Spectres MS/MS des deux toxines réduites et tags obtenus pour une tolérance de 0,02 Da – (A) Toxine à 1328,5142 Da et (B) 1358,5592 Da avant réduction

La présence de trois cystéines n'est pas inattendue puisque l'on sait que chaque peptide en contient au moins quatre. Le premier tag Cys-Cys est retrouvé dans les deux spectres. Le motif formé par deux cystéines vicinales est rencontré dans de nombreuses superfamilles de conotoxines (O-, M-, A-, T- et I-). Le deuxième tag est différent pour les deux toxines. En effet, pour l'ion m/z 1333,5497, le tag Leu(Ile)-Thr-**Pro**-Cys est proposé contre Leu(Ile)-Thr-**Leu(Ile)**-Cys pour l'ion m/z 1363,6038. Ce qui est particulièrement intéressant est que la seule différence observée entre ces deux expériences est qu'une proline (M=97,0528 Da) trouvée dans un cas est remplacée par une leucine ou isoleucine (M=113,0841 Da) dans l'autre cas. La différence de masse entre ces deux acides aminés est de 16 Da, une différence qui a déjà été observée précédemment. A partir de tous ces résultats, on peut donc supposer que cette substitution est la première des deux différences entre les deux toxines.

L'analyse du spectre MS/MS de l'ion à m/z 1363,6038 indique que le fragment à m/z 1233,4900 (perte de 130,1138 Da) est accompagné d'un second ion à m/z 1205,4409, qui correspond à une perte de 28 Da attribuable à la perte d'une molécule neutre de CO, caractéristique des ions de type *b*. Dans cette hypothèse la perte de 130,1138 Da peut être expliquée par la présence d'une leucine (ou d'une isoleucine) amidée positionnée sur l'extrémité C-terminale. L'amidation C-terminale est une modification post-traductionnelle couramment observée chez les conotoxines. La grande précision en masse de l'instrument est très utile puisqu'elle permet de trancher entre la perte d'une Leu/Ile amidée (130,1106 Da) et la perte éventuelle d'une hydroxyproline (130,0742 Da).

Sur le spectre MS/MS de la seconde toxine à m/z 1333,5497, un fragment similaire est observé à m/z 1217,4546. Cette perte de 116,0951 Da est attribuée sans aucune autre possibilité à la perte d'une valine amidée en position C-terminale. Il s'agit donc de la seconde différence entre les séquences des deux toxines, la différence de masse entre une leucine (113,0841 Da) et une valine (99,0684 Da) étant de 14 Da. Aucun autre ion de type *b* n'a été détecté par la recherche de la perte de 28 Da dans ce spectre.

Les autres ions sont donc *a priori* des ions de type *y*, observation confirmée par la présence d'ions satellites correspondant à une perte de 17 Da pour un grand nombre d'entre eux. Cette perte caractérise une perte d'ammoniac (NH₃) couramment observée pour les ions type *y*. Malgré ces informations, le séquençage n'est pas complet et le manque de fragments dans le spectre ne permet pas de conclure sur la séquence du peptide.

Nous avons alors utilisé la précision en masse de l'instrument (environ 10 ppm pour ce type d'expériences) comme un réel outil pour déterminer la composition en acides aminés des « trous » de séquence. Cette étude a été effectuée en prenant en compte les différentes modifications post-traductionnelles rencontrées chez les conotoxines.

B.1.b. Séquençage de la toxine à m/z 1363,6038

La *Tableau 1* regroupe les propositions de *tags* d'acides aminés pour les différentes masses mesurées dans le spectre MS/MS de la toxine à m/z 1363,6038. Un calcul d'erreur exprimé en partie par million (ppm) permet d'évaluer la validité des différentes propositions.

Masses mesurées	Extrémité	Masses recherchées	Possibilités	Masses théoriques	Erreurs (ppm)
226,0952	aucune	226,0952	LL	226,1681	322,4
			OL	226,1518	250,1
			EP	226,0954	0,7
287,2182	C-terminal (acide)	268,2004	HM	268,0994	376,6
			GGGP/GNP	268,1171	310,4
	N-terminal	286,2109	AGGT/GAAS/ASQ/GTQ/ATN	286,1277	290,9
			ASK/GTK	286,1641	163,5
	C-terminal (amidé)	269,1844	GPD	269,1011	309,5
			OGV/APT	269,1375	174,2
			OR	269,1487	132,6
			OR	269,1488	132,2
			AVV/GVL	269,1739	38,8
			LR	269,1852	2,9
214,0426	Nter	214,0426	TL	214,1317	416,3
			OT/VD	214,0953	246,2
			CZ	214,0412	6,5

Tableau 1: Recherche de tags de séquence pour l'ion m/z 1363,6038 par la précision en masse - O et Z désignent respectivement une hydroxyproline et un acide pyroglutamique.

- Recherche de séquence pour la masse de 226,0952 Da

La masse de 226,0952 Da correspond à la différence entre les ions m/z 493,2369 et m/z 719,3321. Etant donné que les deux ions sont de type y , il suffit de comparer directement la différence de masse obtenue aux différentes possibilités. La seule possibilité satisfaisante est l'enchaînement Glu-Pro avec une erreur de 0,7 ppm, excellente pour ce type d'expérience. Les autres possibilités dépassent 200 ppm et sont de ce fait éliminées. Il est utile de préciser que cette recherche ne permet pas de distinguer Glu-Pro de Pro-Glu.

- Etude de l'ion à m/z 287,2182

Pour expliquer la formation de cet ion, nous avons testé toutes les possibilités, à savoir que cet ion peut contenir l'extrémité N-terminale (ion b), l'extrémité C-terminale (ion y) ou l'extrémité C-terminale amidée. De nouveau, les résultats obtenus ne présentent aucune ambiguïté puisque seul le tag Leu-Arg amidé en C-terminal propose une différence de masse adéquate (2,9 ppm). Les autres possibilités conduisent à des erreurs de 40 à 300 ppm. Dans ce cas précis, il est possible de différencier le tag X-Arg-Leu-NH₂ de X-Leu-Arg-NH₂. Néanmoins, lors de l'étude générale des spectres MS/MS nous avons supposé que la présence d'un ion à m/z 1233,4900 s'expliquait parfaitement si une leucine amidée était présente en

position C-terminale. Cette observation permet donc de conclure que le *tag* de séquence est X-Arg-Leu-NH₂.

- Etude de la perte de 214,0426 Da à partir de l'ion parent à m/z 1363,6038

Enfin, nous nous sommes intéressés à la perte de neutre observée entre l'ion parent m/z 1363,6038 et l'ion m/z 1149,5593, qui doit être caractéristique de l'extrémité N-terminale. Les recherches avec les acides aminés classiques ne permettent pas de conclure. Néanmoins, une information supplémentaire est disponible puisque sur les 4 cystéines présentes dans la séquence, seules trois ont d'ores et déjà été localisées. Cette perte doit donc obligatoirement contenir la dernière cystéine. Après avoir retranché la masse de la cystéine aux 214,0426 Da, la masse restant est 111,0334 Da. Cette masse ne peut s'expliquer que par la présence d'un acide pyroglutamique en position N-terminale, modification courante au sein des toxines. Cette modification N-terminale résultant de la cyclisation d'une glutamine (perte de NH₃) explique également les échecs obtenus lors des expériences de dégradation d'Edman tentées sur ces toxines.

Les expériences MS/MS permettent donc de proposer comme séquence pour la première toxine Pyr-Cys-Cys-Leu/Ile-(Glu+Pro)-Cys-Cys-Arg-Leu/Ile-NH₂. L'agencement des cystéines Cys-Cys-(Xxx)_n-Cys-Cys de type V permet de définir la famille d'appartenance de cette conotoxine comme étant la superfamille T. En suivant la nomenclature des conotoxines, cette toxine est appelée conotoxine ViVA, Vi caractérisant *Conus virgo*, V déterminant l'arrangement des cystéines et A car il s'agit de la première toxine de ce type identifiée dans le venin de *Conus virgo*.

B.1.c. Séquençage de la toxine m/z 1333,6038

Une étude similaire a été menée sur la seconde toxine et a conduit aux résultats reportés dans le *Tableau 2*.

Masses mesurées	Extrémité	Masses recherchées	Possibilités	Masses théoriques	Erreurs (ppm)
226,0952	--	226,0955	LL	226,1681	321,1
			OL	226,1518	248,9
			EP	226,0954	0,7
273,2027	C ter (acide)	254,1849	-	-	-
	N ter	272,1954	GGGT/GGAS/GSQ/ASN/GTN	272,1121	306,3
			GSK	272,1485	172,6
	C ter (amidé)	255,1689	GPT/ASP/AAO	255,1219	184,1
			GVV/AAL	255,1583	41,5
			VR	255,1695	2,5
214,0372	N ter	213,0299	GR	213,1226	434,7
			GGV/AAA/VN	213,1113	382,1
			CZ	213,0334	16,3

Tableau 2 : Recherche de tags de la séquence pour l'ion m/z 1333,5797 par la précision en masse -O et Z désignent respectivement une hydroxyproline et un acide pyroglutamique

En suivant la même démarche que précédemment, la séquence proposée pour cette toxine est Pyr-Cys-Cys-Pro-Thr-Leu-(Glu+Pro)-Cys-Cys-Arg-Val-NH₂. L'erreur sur la mesure de masse pour le tag Pyr-Cys (calculée à 16,3 ppm) semble un peu haute mais reste encore tolérable pour des expériences de fragmentation. Cette extrémité est également validée par le fait que les autres possibilités sont inacceptables (>380 ppm) et par la forte homologie présente entre les deux spectres de fragmentation des deux conotoxines.

Les deux seules différences trouvées entre les deux toxines sont bien celles qui avaient été proposées au début de cette étude et permettent d'expliquer les différences de +14, +16 et +30 Da observées entre les deux spectres.

B.2. Levée de l'ambiguïté Glu-Pro ou Pro-Glu

Outre l'ambiguïté leucine/isoleucine habituellement rencontrée dans les expériences de séquençage *de novo*, une seconde inconnue est présente dans les séquences proposées. Il s'agit la localisation précise de la proline et de l'acide glutamique vicinaux. Cette détermination ne peut pas s'effectuer directement sur la base des fragments obtenus en MS/MS. Cependant, la forte intensité des ions à m/z 479,2208 et 493,2369 observée sur les deux spectres et correspondant aux ions y_4 peut être plus facilement interprétée si l'enchaînement du peptide est Pro-Glu-Cys-Cys et pas Glu-Pro-Cys-Cys. Deux raisons peuvent être avancées. La première est liée au clivage favorisé en C-terminal des acides aminés acides. Ce phénomène est amplifié dans le cas d'espèces monochargées dans

lesquelles la charge reste fixée sur un acide aminé basique⁸, ce qui correspond à nos conditions expérimentales. La seconde raison repose sur le fait que des fragmentations intenses du côté *N*-terminal des prolines sont souvent observées⁹. La combinaison de ces deux observations permet d'expliquer les fragments obtenus dans le spectre : absence de fragmentation entre Pro et Glu, et ion y_4 très intense.

Afin de confirmer ces hypothèses, les deux peptides ZCCPTL**PE**CCRV-NH₂ et ZCCPTL**EP**CCRV-NH₂ ont été synthétisés et leurs spectres de fragmentation ont été enregistrés dans les mêmes conditions que celles utilisées pour les toxines natives (*Figure 7*). Les peptides ont été synthétisés **arbitrairement** avec des leucines.

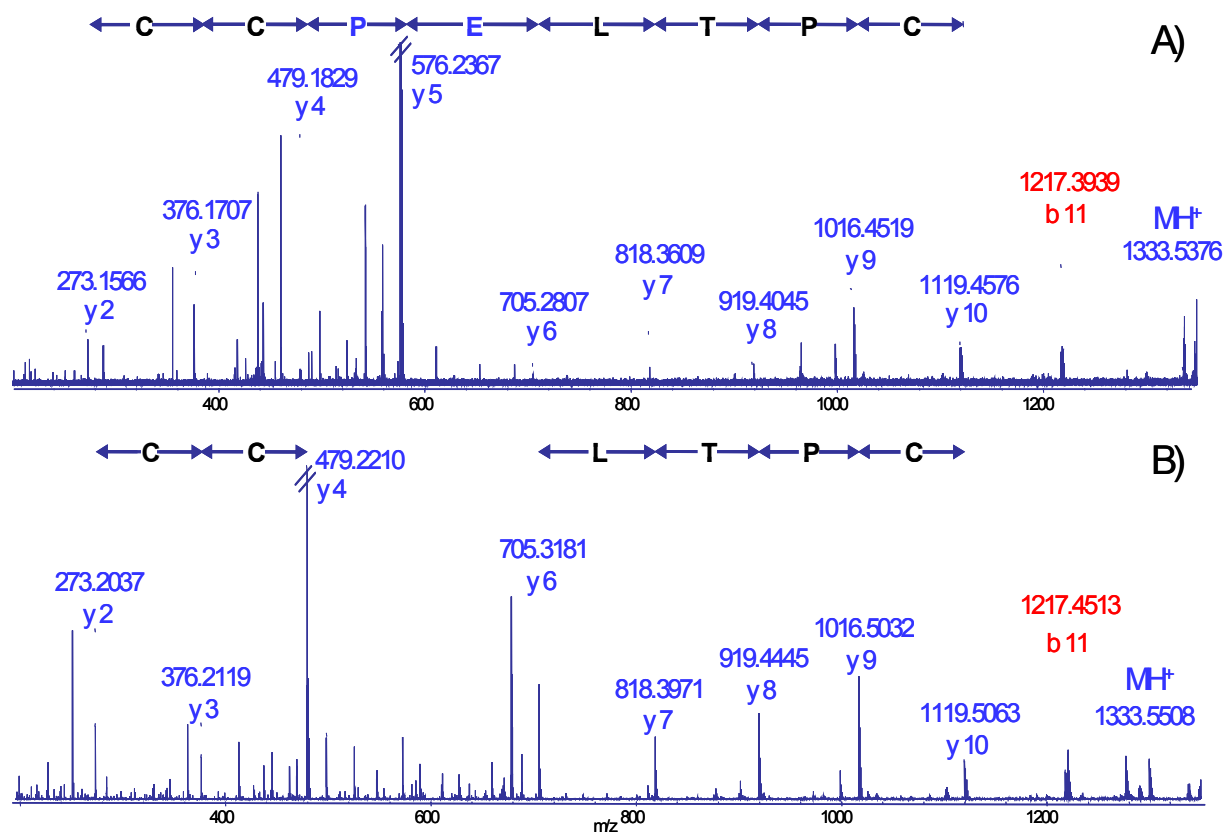


Figure 7: Spectres MS/MS des toxines de synthèses (A) ZCCPELTPCCRV-NH₂ et (B) ZCCEPLTPCCRV-NH₂

Les deux spectres obtenus présentent des fragmentations similaires dues aux homologies entre les deux séquences. Néanmoins, deux différences majeures peuvent être observées. La première est que sur le spectre (A), l'ion y_5 est le plus intense alors qu'il n'est pas du tout observé sur le spectre (B). De plus, l'ion y_4 , très intense sur le spectre (B) est très faiblement observé sur le (A). La comparaison de ce spectre avec les résultats obtenus (*Figure 6*) montre

sans ambiguïté que nos toxines natives possèdent bien un enchaînement -Pro-Glu-. Les résultats obtenus sont résumés dans le *Tableau 3*.

Séquence de la toxine	Pyr-C-C-I/L-T-I/L-P-E-C-C-R-I/L-NH ₂	Pyr-C-C-P-T-I/L-P-E-C-C-R-V-NH ₂
Masse théorique (Da)	1358,5604	1328,5134
Masse expérimentale (Da)	1358,5592	1328,5142
Erreur (ppm)	0,9	0,6
Ions fragments observés	<p>The diagram shows the peptide sequence Pyr-C-C-I/L-T-I/L-P-E-C-C-R-I/L-NH₂. Above the sequence, y-ion fragments are indicated by arrows pointing to specific residues: y₁₀ (C), y₉ (I/L), y₈ (T), y₇ (I/L), y₆ (P), y₅ (E), y₄ (C), y₃ (C), and y₂ (R). Below the sequence, the b₁₁ ion is indicated by an arrow pointing to the C-terminal residue (R).</p>	<p>The diagram shows the peptide sequence Pyr-C-C-P-T-I/L-P-E-C-C-R-V-NH₂. Above the sequence, y-ion fragments are indicated by arrows pointing to specific residues: y₁₀ (C), y₉ (P), y₈ (T), y₇ (I/L), y₆ (P), y₅ (E), y₄ (C), y₃ (C), and y₂ (R). Below the sequence, the b₁₁ ion is indicated by an arrow pointing to the C-terminal residue (R).</p>

Tableau 3: Récapitulatif des résultats obtenus sur le venin de *Conus virgo*

La présence d'une arginine proche de l'extrémité C-terminale permet d'expliquer que seule la série d'ions *y* est observée. En effet, puisque les expériences de MS/MS ont été effectuées sur l'espèce $[M+H]^+$, l'unique proton du système est séquestré par l'arginine, résidu très basique. Les ions *y* qui la contiennent sont donc majoritairement observés. Le seul ion *b* observé dans le spectre est *b*₁₁. C'est également le seul ion *b* qui contienne l'arginine.

Des expériences de réduction ménagées ont été menées, en vain, pour déterminer le type de branchement des ponts disulfures de ces toxines. Néanmoins, le motif formé par les ponts disulfures étant conservé au sein de la superfamille-T, on peut supposer que ce même branchement est observé ici, à savoir, Cys₂-Cys₉ et Cys₃-Cys₁₀.

Les recherches d'activité biologique de ces deux toxines (par le Dr Nicolas Gilles CEA, Saclay) n'ont pour l'instant donné aucun résultat mais de nombreuses expériences restent encore à effectuer. Quoi qu'il en soit cette étude a mis en évidence l'existence de deux nouvelles conotoxines T, qui constituent de nouveaux outils disponibles pour avancer dans la compréhension du rôle biologique de cette famille.

Nos résultats montrent que notre approche directe par nanoESI-FT-ICR sur le venin brut permet d'établir l'empreinte massique de venin et de séquencer deux peptides directement dans le mélange brut. Cependant, la large gamme de masse observée et le peu de composés finalement détectés par rapport à ce que l'on sait de la richesse habituelle des

venins de cônes nous conduit à supposer que l'approche directe ne permet de détecter qu'une partie des constituants. Cette discrimination entre différentes familles de composés possédant des caractéristiques physico-chimiques distinctes est par ailleurs sans doute accentuée par le phénomène de suppression ionique due à l'utilisation du nano-electrospray comme technique d'ionisation. Bien qu'il soit établi que ce problème soit moins crucial en nano-electrospray qu'en electrospray classique^{10,11}, une comparaison des résultats obtenus par notre approche directe avec ceux obtenus par l'utilisation du couplage nanoHPLC-FT-MS a été effectuée.

C. Apports de l'approche par couplage en ligne

C.1. Intérêt du couplage nanochromatographie/spectrométrie de masse pour l'analyse de mélanges complexes

L'intérêt de l'utilisation de la nanochromatographie en phase liquide réside dans la possibilité d'atteindre des efficacités de séparation très élevées et de travailler avec des volumes d'échantillon très faibles (et donc de faibles quantités de matériel). Par exemple, une centaine de femtomoles d'un digestat protéique suffisent pour obtenir une cartographie peptidique exploitable. De plus, les composés sont élués tour à tour de la colonne analytique, ce qui permet d'augmenter la gamme dynamique. Enfin, une microcolonne placée en amont de la colonne analytique permet de dessaler et de concentrer les échantillons en ligne. Ce montage permet d'améliorer la qualité des analyses en éliminant les contaminants comme les sels (ou autres produits trop polaires pour se fixer sur la colonne) et d'éviter des pollutions éventuelles de la colonne analytique et du spectromètre de masse dans le cas du couplage en ligne « direct ». Néanmoins, nous montrerons que cette précolonne peut également être responsable de la perte de composés.

Jusqu'à présent aucune publication (source PubMed) ne rapporte l'analyse de venins par nanochromatographie. La plupart des études s'effectuent en général par fractionnement par chromatographique du venin avant l'analyse par spectrométrie de masse (ESI, nanoESI, MALDI). Le couplage entre une chaîne de chromatographie liquide à nanodébit et notre FT-ICR a été réalisé pour la première fois au cours de cette thèse. Les premiers essais ont été effectués sur des digestats de protéines. Nous avons ensuite optimisé les conditions expérimentales (le gradient) notamment, en prenant en considération le fait qu'un venin brut contient des constituants aux propriétés bien plus variées qu'un digestat.

C.2. Analyse du venin brut par nanoLC-FT-ICR

Le chromatogramme obtenu pour l'analyse en nanoLC-FT-ICR du venin de *Conus virgo* est présenté sur la Figure 8.

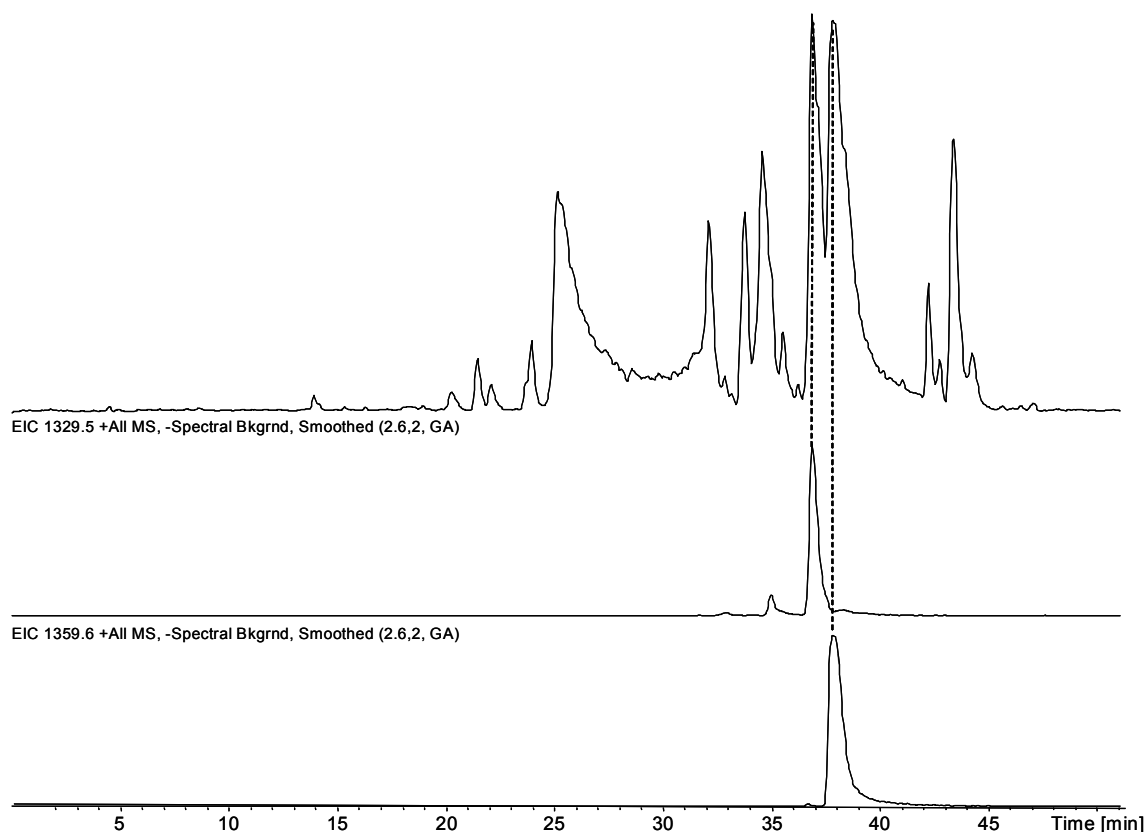


Figure 8: Chromatogramme (reconstruit sur l'intensité du pic de base) obtenu pour le venin de *Conus virgo* en nanoLC-FT-ICR – En dessous, sont extraits les chromatogrammes de (B) ViVA et (C) ViVB

L'analyse de ce chromatogramme révèle la présence de 130 masses moléculaires différentes qui appartiennent à la gamme de masse [500-7000] Da, plus étendue que celle observée lors de l'analyse directe par nanoESI (500-4500 Da). Les traces chromatographiques des deux toxines ViVA et ViVB séquencées dans le venin brut sont reportées sous le chromatogramme total. Il apparaît nettement que ces deux toxines, majoritaires pour les expériences en infusion directe du venin brut, correspondent également à des pics majoritaires dans le chromatogramme.

C.3. Comparaison infusion directe/couplage pour le venin de *Conus virgo*

Le premier point à noter est que, comme attendu, le couplage permet de détecter un plus grand nombre de composés que l'infusion directe. Néanmoins, une observation surprenante est que sur les 130 masses moléculaires mesurées en couplage, seulement 45 sont communes avec celles issues des expériences en infusion directe. Ces 45 masses communes représentent environ 90% du signal total du spectre nanoESI-FT-ICR et 71% de l'intensité des pics obtenus en nanoLC-FT-ICR. Ceci montre clairement que les 21 composés perdus et les 66 gagnés par l'utilisation du couplage correspondent à des espèces de faible abondance.

La « disparition » de composés présents dans un venin brut lors de son analyse en couplage a déjà été observée par Taggi *et al.*¹² alors qu'ils cherchaient à caractériser des nucléosides dans un venin brut d'araignée. Leur conclusion est que les conditions expérimentales de la chromatographie jouent un rôle crucial sur le nombre de composés qu'il est possible de retrouver après les différentes étapes de la séparation. L'analyse du venin brut sans aucune séparation préalable permet de s'affranchir de ce problème et peut donc apparaître comme complémentaire d'une approche par couplage.

Afin de déterminer s'il existe une corrélation entre la masse moléculaire des composés et leur perte lors de l'étape de séparation, nous avons dénombré les composés observés (i) en infusion directe (ii) en couplage et (iii) communs aux deux approches, pour différentes gammes de masse.

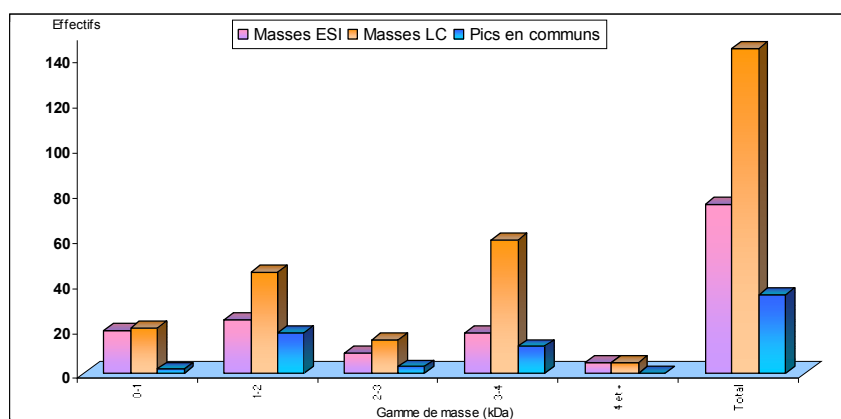


Figure 9: Comparaison du nombre de masse détectées pour les deux types d'expériences – en violet pour le nanoESI, en vert pour la nanoLC-MS, en bleu le nombre commun aux deux expériences

L'analyse de la *Figure 9*, qui regroupe ces résultats, apporte différentes informations. Tout d'abord, le nombre de masses détectées à l'issue du couplage est toujours supérieur (quelle que soit la gamme de masse) à celui obtenu en infusion directe, validant l'approche par couplage pour détecter un maximum de composés dans un venin. La deuxième information est que la perte de matériel due au couplage semble répartie sur toute la gamme de masse. En d'autres termes, le couplage n'entraîne pas particulièrement la disparition de composés de faible poids moléculaire ou au contraire de poids moléculaire élevé.

Pour essayer de comprendre ce phénomène, la comparaison de l'approche directe nanoESI et du couplage nanoLC-FT-ICR a été effectuée sur une digestion trypsique de myoglobine de coeur de bœuf. Les résultats de cette étude permettent de tirer des conclusions similaires à ce qui est observé pour le venin puisque plusieurs peptides tryptiques détectés dans les expériences de nanoESI-FT-ICR ne sont pas retrouvés après la séparation chromatographique. Des expériences complémentaires en ajoutant un détecteur UV après la précolonne mettent en évidence que certains peptides ne sont pas retenus par la précolonne de dessalage utilisée et sont perdus lors du rinçage de cette dernière. Des améliorations doivent donc être apportées à notre système (changement du type de précolonne, optimisation des conditions de dessalage des échantillons etc...) afin qu'il soit réellement performant pour l'étude des venins bruts, sans perte importante de matériel.

D. Conclusion

Nous avons montré lors de l'étude de *Conus virgo* que l'analyse directe, par spectrométrie de masse FT-ICR, d'un venin brut était tout à fait possible, si l'on accepte de se limiter à la caractérisation des composés majoritaires. Pour une étude plus exhaustive, l'utilisation d'une méthode séparative avant la spectrométrie de masse est nécessaire. De la même façon, cette séparation est nécessaire si le mélange est extrêmement complexe car une approche directe conduit à une perte d'informations importante.

III. Analyse du venin de *Conus ermineus* : comparaison de différentes approches

Conus ermineus est un cône qui provient des côtes sénégalaises et qui mesure en moyenne 5 cm. Ce cône a jusqu'alors été très peu étudié mais des études menées au CEA de Saclay ont permis de séquencer une nouvelle δ -conotoxine nommée δ -EVIA^{13,14}. Cette toxine lie sélectivement les canaux sodiques neuronaux et, en association avec un marqueur approprié, permet de les localiser au niveau cellulaire et d'en étudier la distribution. Cette activité permet d'envisager de pouvoir l'utiliser dans le traitement symptomatique de différentes maladies neurodégénératives du système nerveux comme la sclérose en plaques. Cette récente découverte illustre parfaitement tout l'intérêt qui réside dans l'exploration détaillée des venins de cônes.

Plusieurs difficultés liées à l'analyse de ce venin ont orienté notre stratégie analytique. La plus importante est que nous ne disposions pas d'une quantité importante de venin brut puisqu'environ 3 mg seulement ont été obtenus et lyophilisés. Notre objectif était de déterminer un nombre maximum de candidats intéressants d'un point de vue biologique avec un minimum d'échantillon. La première étape pour réaliser cette sélection est d'effectuer une cartographie massique du venin brut avant et après réduction de tous les ponts disulfures. À l'issue de ces expériences, deux types d'informations sont obtenus: la masse précise des toxines ainsi que le nombre de ponts disulfures. Comme nous l'avons vu précédemment, le nombre de ponts disulfures présent dans une conotoxine est une information importante puisqu'elle permet de classer plus ou moins les composés dans les différentes familles de conotoxines.

La cartographie massique du venin de *Conus ermineus* avant et après réduction a été envisagée dans un premier temps par nanoESI directe. Cependant, le nombre important de composés détecté nous a conduit à introduire une étape de séparation pour effectuer les expériences dans les meilleures conditions possibles.

A. Cartographie massique de *Conus ermineus* en infusion directe

Bien que l'échantillon ait été purifié sur microcolonne de type ZipTip[®] le spectre résultant (*Figure 10*) est particulièrement difficile à interpréter.

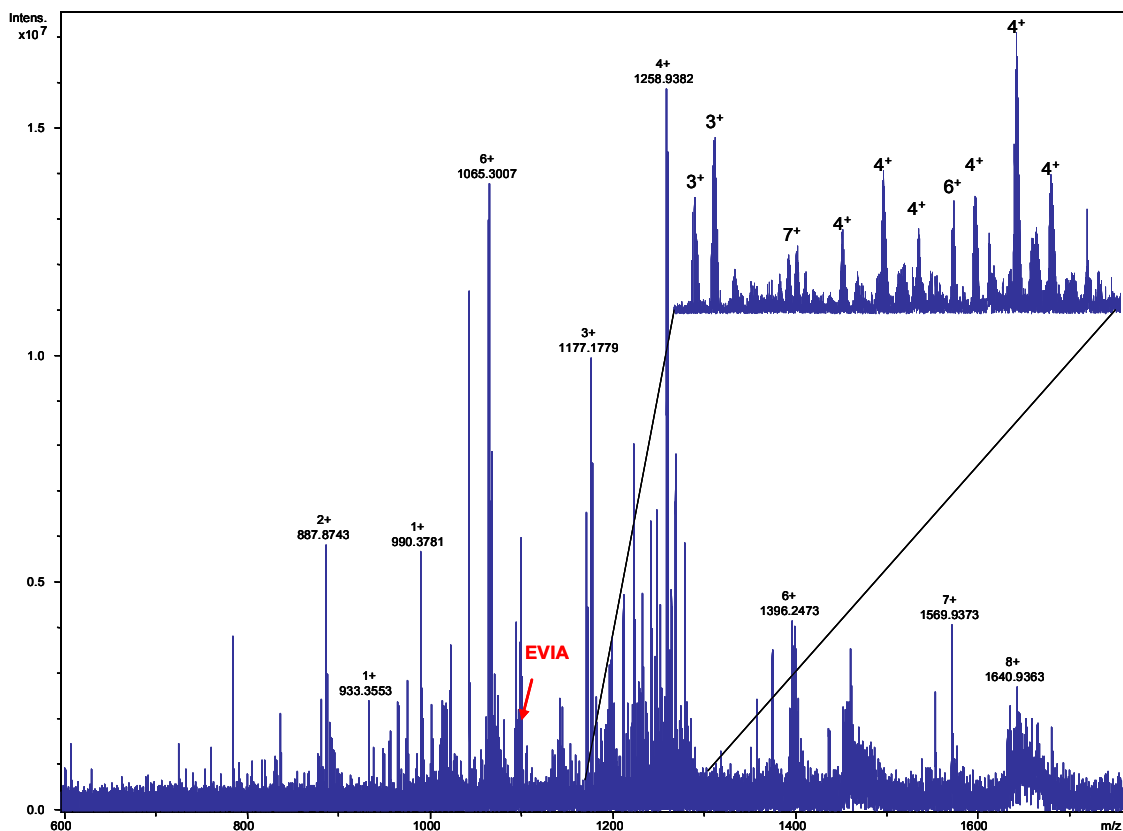


Figure 10: Spectre nESI-FT-ICR du venin brut de *Conus ermineus* – 50 scans

Grâce à la résolution du FT-ICR, on peut détecter plus de 70 composés appartenant à la gamme de masse [500-13500] Da. La δ -conotoxine EVIA ($M_r=3285,4459$ Da) déjà étudiée par le CEA de Saclay est détectée avec une très faible intensité de signal.

Les toxines détectées dans le venin lors de cette étude présentent une distribution relativement homogène sur une gamme de masse allant de 0 à 9 kDa (*Figure 11*). On observe néanmoins un nombre plus important de composés compris entre 3 et 4 kDa représentant 27% de l'intensité totale observée.

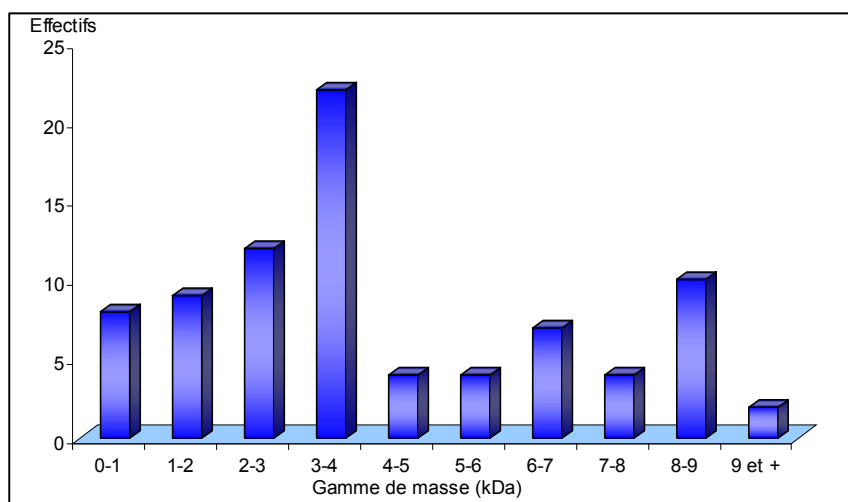


Figure 11: Répartition des masses des composés détectés dans le venin de *Conus ermineus* analysé par nanoESI-FT-ICR

Le diagramme montre la présence d'un nombre important de composés dans la gamme de masse des δ -conotoxines (3000-4000 Da). La deuxième étape a été de réduire tous les ponts disulfures afin de corrélérer les masses observées avec un nombre de ponts S-S. Malheureusement, la complexité du spectre après réduction (recouvrement de nombreux massifs isotopiques) rend totalement impossible son exploitation. Les deux cartographies avant/après réduction ont donc été effectuées en utilisant le couplage nanoLC-FT-ICR.

B. Approche par un couplage en ligne nanoLC-FT-ICR

Compte-tenu du peu d'échantillon disponible, l'utilisation d'une chaîne de chromatographie à nanodébit est indispensable. Comme explicité plus haut, le but de ces expériences (résumées sur la *Figure 12*) était de pouvoir cibler plus ou moins les toxines intéressantes en fonction non seulement de leur masse mais aussi du nombre de ponts S-S.

D'un point de vue technique, les expériences de couplage en ligne entre la nanoLC et le FT-ICR ne permettent pas d'utiliser la très haute résolution de la FT-ICR en raison de la taille des fichiers de données obtenus qui seraient alors trop lourds pour être retraités par le système informatique. A titre d'exemple, la taille du fichier comprenant un chromatogramme enregistré sur une période d'une heure (1400 spectres) en résolution moyenne (256 000 points/spectres) dépasse facilement un gigaoctet. Ceci représente une contrainte non

seulement pour le retraitement puisqu'il n'est pas aisé de manipuler des fichiers de cette taille mais aussi pour le stockage.

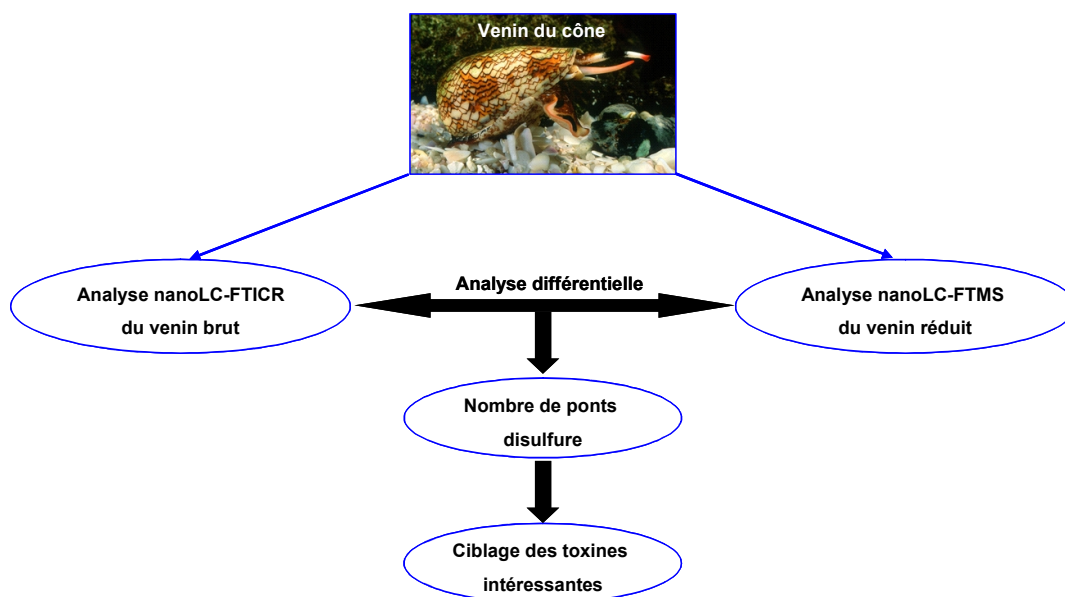


Figure 12: Stratégie d'analyse employée pour cibler les toxines d'intérêt présentes dans le venin de *Conus ermineus*

B.1. Analyse nanoLC-FT-ICR du venin avant réduction

L'analyse du chromatogramme obtenu pour l'analyse en couplage du venin de *Conus ermineus* (Figure 13) permet de se rendre bien mieux compte de sa complexité. De nombreux pics très fins sont détectés, montrant l'efficacité de notre système et la qualité de la séparation. Chacun des différents pics chromatographiques est caractéristique d'une ou de plusieurs molécules. Au total, ce sont plus de 370 composés différents qui sont détectés dans une gamme de masse très large : [500 Da - 25 000 Da].

Une étude plus fine de la distribution massique des différents composés (Figure 14) montre une hétérogénéité sur l'ensemble de la gamme de masse. En effet, une majorité des composés présente une masse comprise entre 1000 et 5000 Da (80% environ), même si un nombre non négligeable d'entre eux est mesuré au-delà. Cette distribution est typique des venins de cônes marins qui possèdent généralement de nombreux peptides avec des masses inférieures à 5000 Da.

Chromatogramme

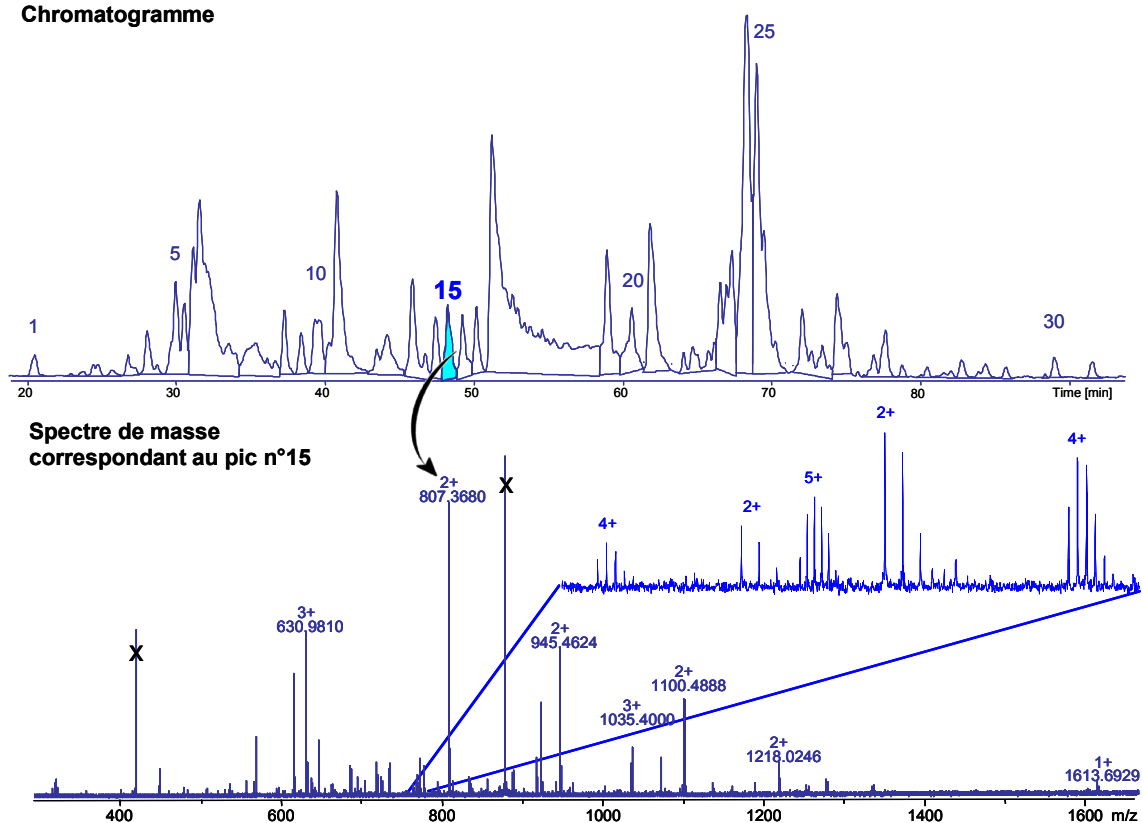


Figure 13: L'analyse nanoLC-FT-ICR en ligne du venin de *Conus ermineus* révèle environ 370 masse moléculaires différentes

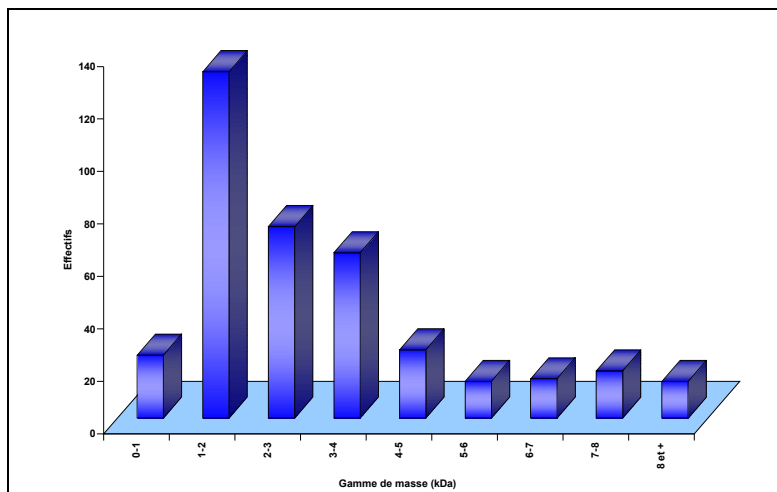


Figure 14: Distribution massique des composés détectés dans le venin brut de *Conus ermineus* par nanoLC-FT-ICR

Il est à noter que, dans cette étude, la chromatographie est utilisée simplement comme une technique séparative et non comme un outil analytique à part entière qui permet de fournir des informations *via* les temps de rétention observés.

B.2. Analyse du venin réduit en nanoLC-FT-ICR

La même analyse a été effectuée sur le venin après réduction des ponts disulfures. Le chromatogramme obtenu est présenté sur la *Figure 15*.

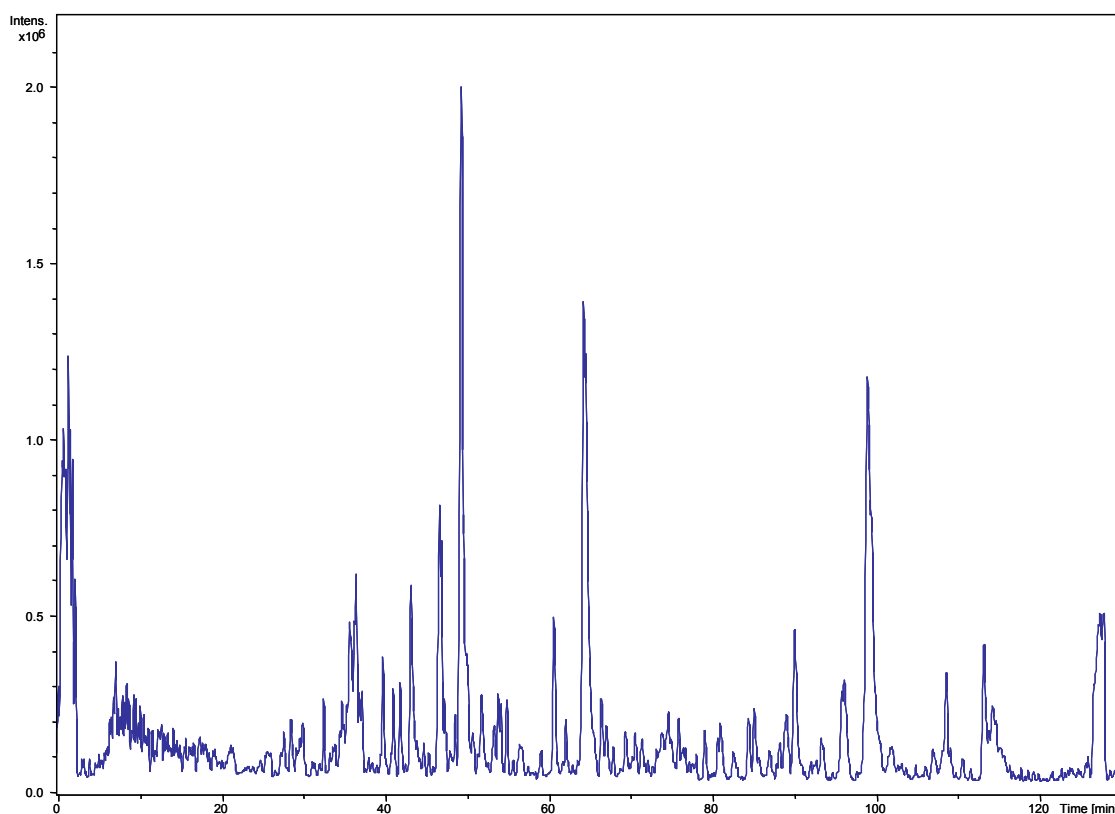


Figure 15: Chromatogramme obtenu pour l'analyse en nanoLC-FT-ICR du venin de *Conus ermineus* après réduction des ponts S-S

L'analyse du chromatogramme permet de déterminer 226 masses moléculaires distinctes entre 500 et 8 000 Da. Ces résultats montrent que de nombreux composés ne sont pas retrouvés puisque 370 avaient été dénombrés avant la réduction. Cette perte importante peut être due à un problème purement chromatographique, ou à un problème de préparation d'échantillon comme, par exemple, une dégradation des composés lors de la réduction des ponts disulfures (50°C pendant 30 min).

La comparaison des deux analyses (avant/après réduction) permet de déterminer le nombre de ponts disulfures pour 110 composés différents, soit seulement un tiers des composés détectés

en nanoLC-FT-ICR avant la réduction. Le fait de perdre autant de composés montre que les conditions du couplage peuvent être améliorées. La répartition des 110 composés en fonction du nombre de ponts disulfures est reportée sur la *Figure 16*.

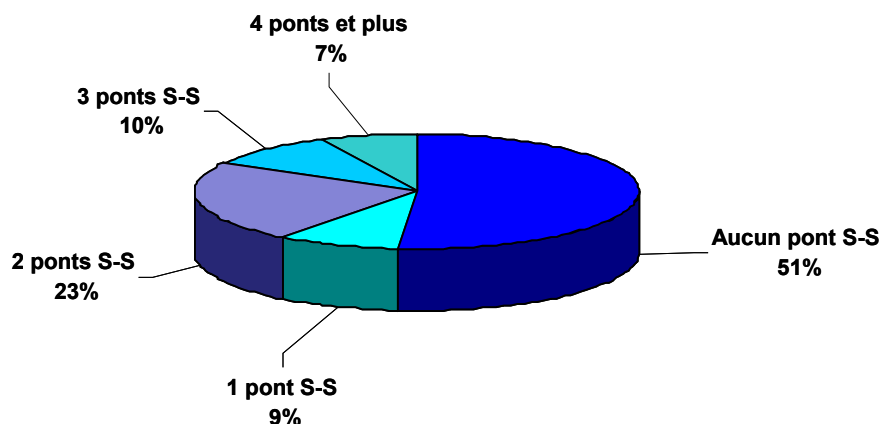


Figure 16: Nombre de ponts disulfures des 110 composés du venin de *Conus ermineus* étudiés

Cette répartition montre que la moitié des composés du venin ne possède pas de pont disulfure ou que ces derniers n'ont pas été réduits. Aucune déformation des massifs isotopiques observés dans les spectres de masse (indiquant une réduction partielle de certains peptides) n'a été observée ce qui pourrait suggérer que cette répartition est bien le reflet de la composition du venin brut. Cependant, étant donné la forte proportion de peptides non réduits, une hypothèse plus vraisemblable serait qu'une partie des conotoxines du venin, plus structurées, n'ait pas été réduites.

- Toxines sans pont disulfure. les conopeptides ne possédant aucun pont disulfure sont ceux appartenant aux familles des contulakines, des conantokines ou des conorfamides.
- Toxines possédant un seul pont disulfure. Les 9% de peptides qui possèdent un unique pont disulfure appartiennent soit à la famille des conopressines actives sur les récepteurs à la vasopressine, soit à celle des contryphans dont les cibles biologiques sont encore mal connues.
- Toxines possédant deux ponts disulfures. Elles représentent 23% des toxines retrouvées et elles appartiennent soit à la superfamille *A* et plus précisément aux α -conotoxines (actives sur les récepteurs à l'acétylcholine), soit à la superfamille *T*, dont les récepteurs cibles ne sont, également, pas encore bien connus.

○ Toxines possédant trois ponts disulfures. La famille des toxines à trois ponts disulfures, très complexe, représente 10% de la totalité des composés détectés. Le fait de déterminer ce nombre de ponts disulfures est relativement peu informatif sur la famille des toxines puisqu'elles peuvent appartenir à une multitude de familles dont les δ -, μ O-, κ - et ω -conotoxines pour la O-superfamily, les μ -, ψ - et κ M-conotoxines pour la M-superfamily et les α A- et κ A-conotoxines pour la A-superfamily.

○ Toxines possédant quatre ponts ou plus Le reste de composés, représentant 7% des masses étudiées possède 4 ponts disulfures ou plus, caractéristique des superfamilles I, active sur les canaux ioniques (potassium) dépendant du potentiel membranaire, et S active sur les récepteurs 5-hydroxytryptamine (5-HT) sensibles à la sérotonine. Enfin, il peut s'agir de protéines diverses comme des enzymes que l'on peut également trouver dans ces venins.

Outre les familles citées, il est également envisageable dans chacun de ces cas que certains de ces peptides appartiennent à une famille encore inconnue.

La *Figure 17* regroupe les résultats obtenus par notre approche et ceux déjà décrits pour les différentes conotoxines trouvées dans la littérature et déjà présentées sur la *Figure 1*.

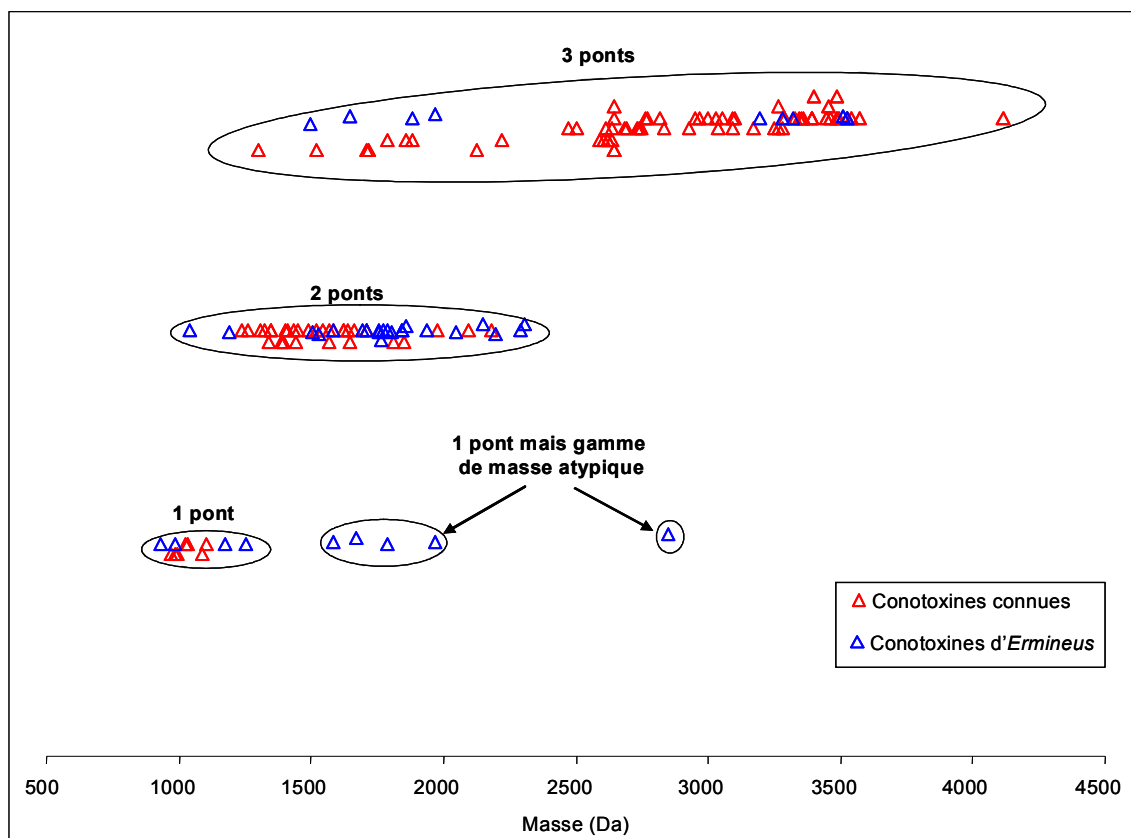


Figure 17: Répartition des masses des conotoxines du venin de *Conus ermineus* en fonction du nombre de ponts disulfures dans leur séquence –Seules les familles possédant 1, 2 ou 3 ponts sont répertoriées ici.

L'analyse de la *Figure 17* montre que les peptides de *Conus ermineus* (en bleu) possédant deux ou trois ponts S-S ne sortent pas des limites, en masse moléculaire, observées pour les conotoxines décrites dans la littérature. En revanche, des résultats beaucoup plus intéressants sont obtenus pour les peptides avec un pont disulfure qui présentent une gamme de masse bien plus élargie (vers les hautes masses) que ce qui avait été observé jusque là. En effet, on retrouve des peptides avec un seul pont disulfure pour des masses allant de 1500 à 3000 Da. Ces résultats sont très intéressants car il pourrait s'agir d'une nouvelle famille de conotoxines. Leur trop faible intensité dans les différents spectres ne permet pas, dans nos conditions expérimentales, d'obtenir des éléments de séquence pour ces composés. Malgré tout, leur découverte valide l'approche que nous avons développée ici, approche qui permet de cibler très rapidement les composés qui possèdent des propriétés différentes de ceux déjà connus.

C. Approche par un couplage hors ligne entre la chromatographie d'échange d'ions et la FT-ICR

Comme nous venons de le voir, l'utilisation du couplage nanoLC-FT-ICR en ligne est performante pour réaliser des cartographies massiques mais l'est beaucoup moins pour obtenir des informations de séquence. Ceci est principalement dû au fait que les expériences de MS/MS en SORI-CAD ou en ECD ne peuvent être réalisées efficacement en un temps court (durée d'un pic chromatographique) et que de nombreuses accumulations sont souvent nécessaires pour obtenir des spectres de fragmentation exploitables. L'absence d'automatisation pour la sélection des ions à fragmenter est un autre problème.

Il faut noter qu'aucune conotoxine n'a jamais été caractérisée complètement par des expériences en LC-MS/MS ou nanoLC-MS/MS. Les seules informations apportées par des études de ce type sont des *tags* de séquence de quelques acides aminés¹⁵.

La stratégie analytique a donc été modifiée. Malgré la faible quantité de matériel, un fractionnement du venin par chromatographie d'échange d'ions a été effectué (*Figure 18*).

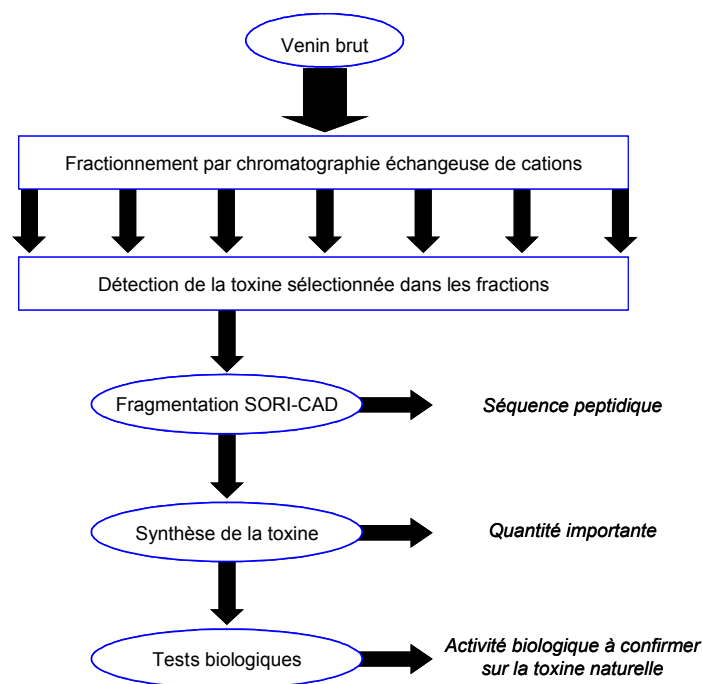


Figure 18: Stratégie adoptée pour caractériser complètement des toxines de *conus ermineus*

Cette étape initiale permet d'obtenir des échantillons moins complexes ce qui est nécessaire pour les expériences MS/MS envisagées. Le fractionnement par chromatographie échangeuse de cations permet de recueillir 11 fractions (Figure 19) nommées de Er₁A à Er₁K (Er : *Ermineus*). Chacune des fractions a ensuite été analysée par nanoESI-FT-ICR (infusion directe) avant et après réduction des ponts S-S.

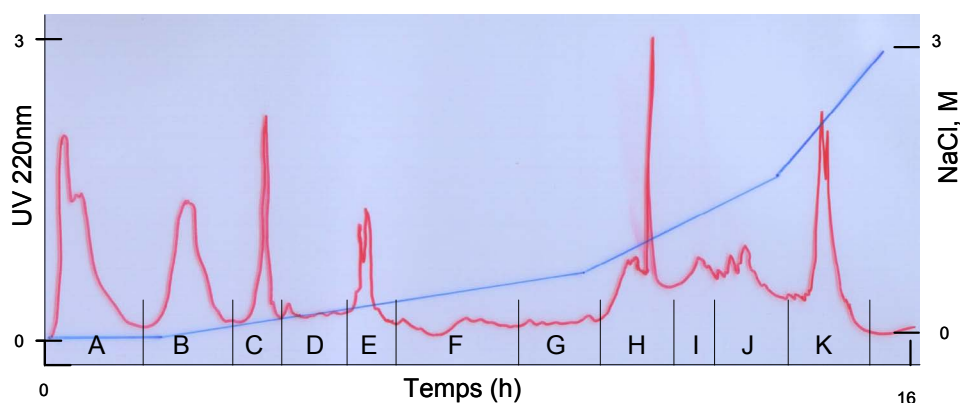


Figure 19 : Chromatogramme de fractionnement du venin de *Conus ermineus* par chromatographie échangeuse de cations

Afin de valider notre stratégie d'analyse, deux peptides présentant un nombre de ponts différents ont été sélectionnés dans des fractions (K et D). Ces deux fractions ne semblent pas contenir la même quantité de matériel d'après la trace chromatographique. Le premier

peptide, observé dans la fraction Er₁K, a une masse de 989,3490 Da. Le second composé (Er₁D), a une masse de 1772,6973 Da. Ces deux composés sont majoritaires dans les fractions K et D, respectivement. Ces deux fractions ont été réduites par du TCEP, purifiées sur microcolonne ZipTip® C₁₈ puis analysées par nanoESI-FT-ICR. Les expériences de fragmentations menées ont été effectuées par collision avec du xénon (SORI-CAD).

C.1. Séquençage *de novo* dans la fraction K

La masse du peptide à 989,3490 Da présent dans la fraction K est déplacée après réduction à 991,3668 Da, caractéristique de la présence d'un seul pont disulfure (*Figure 20*). Pour nommer cette toxine potentielle, le nom de Er₁K₉₈₉ sera utilisé en relation avec le nom du cône étudié, la fraction dans laquelle on l'a retrouvée après purification et sa masse.

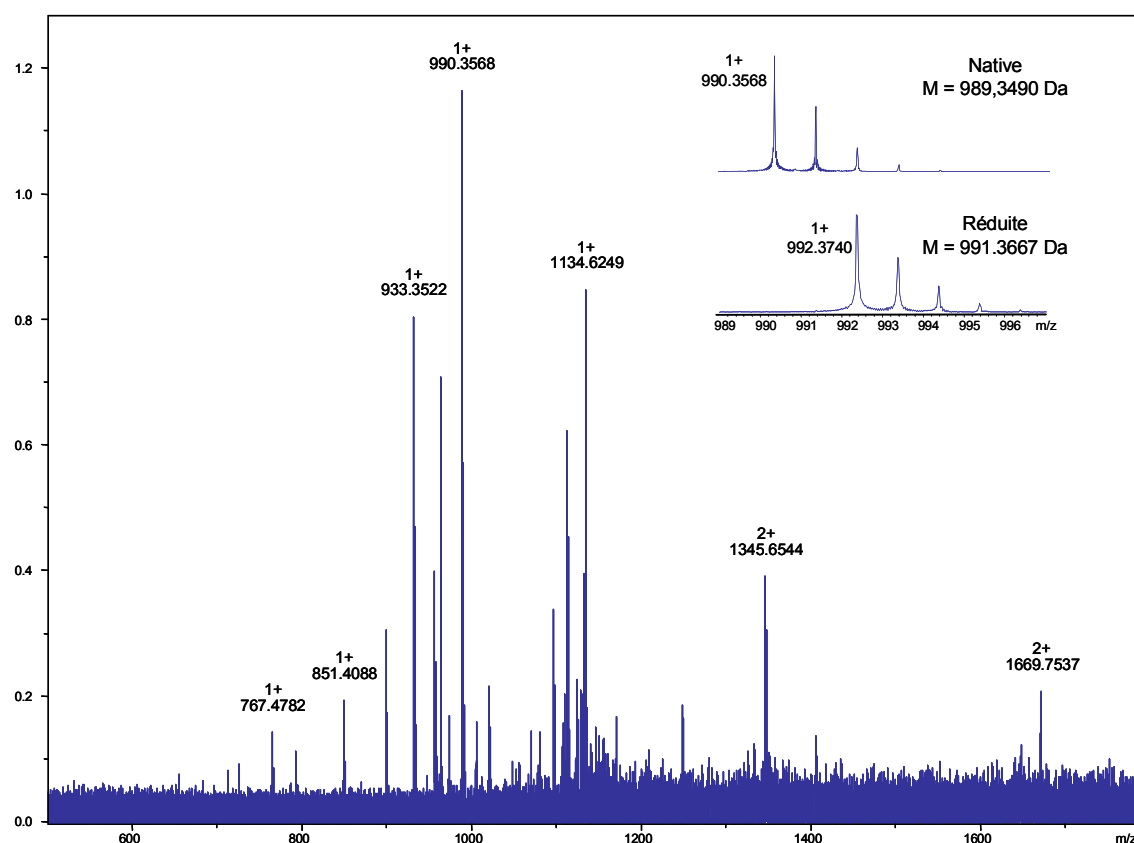


Figure 20: Spectre nanoESI-FT-ICR de la fraction Er₁K – En haut à droite, un zoom du signal correspondant au peptide d'intérêt obtenu avant et après réduction des ponts S-S

Le peptide d'intérêt n'étant détecté que sous la forme d'un ion monochargé, une approche par ECD n'a pu être tentée. La toxine a donc été fragmentée par collision avec du xénon en SORI-CAD (Figure 21).

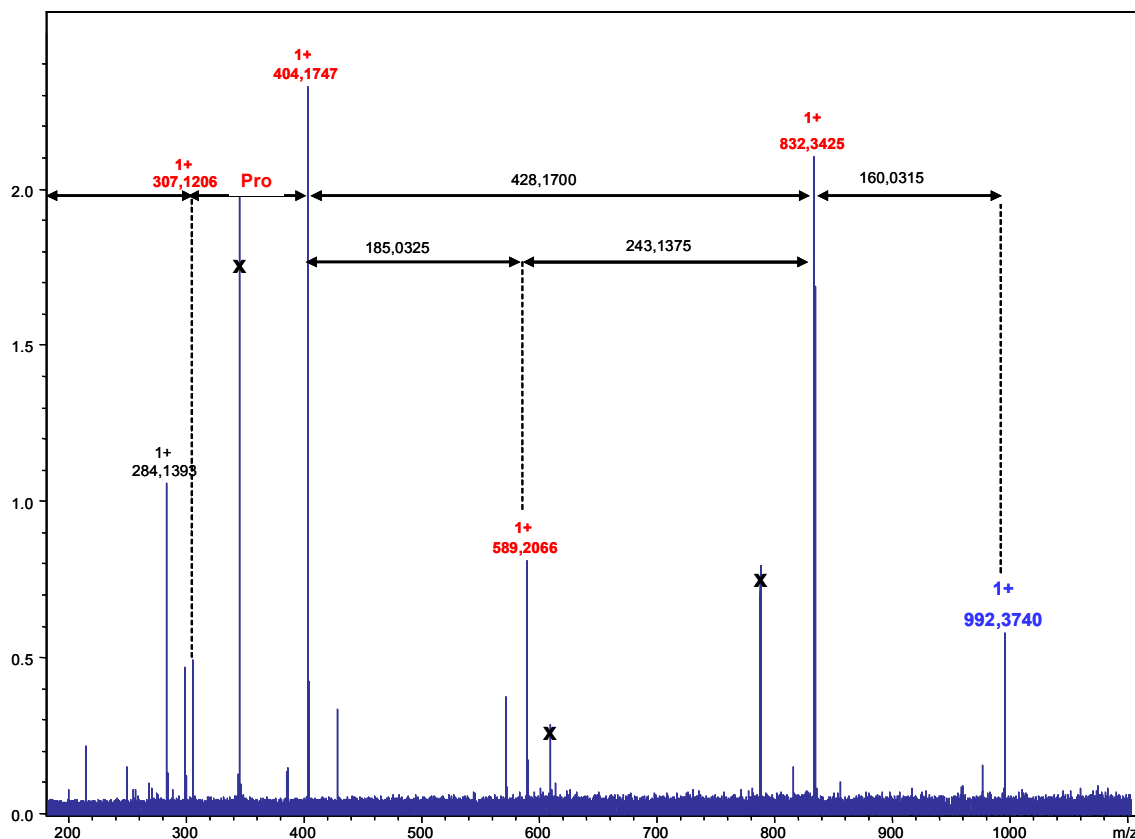


Figure 21: Spectre SORI-CAD (Xe) de la toxine réduite dans la fraction Er₁K₉₈₉

Le spectre MS/MS obtenu contient peu de fragments : on observe majoritairement la présence de trois ions intenses à m/z 832,3425, m/z 589,2066 et m/z 404,1747. L'absence de nombreux fragments est similaire à ce que l'on avait observé pour la fragmentation des toxines de *Conus virgo*, et qui avait été, pour partie, expliquée par la présence de l'enchaînement Glu-Pro dans la séquence. Il est possible que cet enchaînement soit également présent ici puisque la seule information que l'on peut déduire de l'analyse de ce spectre est la présence d'une proline (par différence entre m/z 404,1747 et m/z 307,1206). Outre cet acide aminé, aucune autre information de séquence primaire n'est lisible directement sur le spectre.

Pour aller plus loin dans cette caractérisation, la même technique de recherche de structure que pour *Conus virgo* a été appliquée, c'est-à-dire la comparaison des masses mesurées entre les différents ions majoritaires avec les masses de différentes possibilités de séquence (Tableau 4).

Masses mesurées	Extrémité	Masses recherchées	Possibilités	Masses théoriques	Erreurs (ppm)
307,1206	N ter	306,1133	aucune	--	--
	C ter (acide)	288,1028	GGSS/SSN	288,1070	14,7
			GGAC/GCQ/ACN	288,0892	47,1
			GCK	288,1256	79,3
	C ter (amidé)	289,0868	CW	289,0885	5,9
			SSD	289,0910	14,6
			GCE/ACD	289,0732	47,0
			ASM/GTM	289,1096	78,9
			STT	289,1274	140,5
160,0315	N ter	160,0315	GC	160,0307	5,3
243,1375	aucune	243,1375	SR	243,1331	18,1
			DK/GSV/AAT	243,1219	64,2
			GW	243,1008	151,0
185,0325	aucune	185,0325	GGA/GQ/AN	185,0800	256,6

Tableau 4 : Proposition d'éléments de séquence pour le peptide Er1K₉₈₉ – Les possibilités données dans le tableau correspondent à celles trouvées pour une tolérance de 0,05 Da au maximum -

Pour la première extrémité, caractérisée par l'ion à m/z 307,1206, trois possibilités (*N*-ter et *C*-ter acide ou amide) sont testées. Aucune séquence ne correspond à cette masse s'il s'agit de l'extrémité *N*-terminale (pour une tolérance de recherche de 0,05 Da). Pour un *C*-terminal acide, deux propositions sont trouvées avec un écart de l'ordre de 15 ppm, ce qui semble un peu élevé d'autant que pour une extrémité *C*-terminale amidée, l'enchaînement Cys-Trp conduit à une erreur très satisfaisante de moins de 6 ppm.

Pour l'autre extrémité du peptide, le *N*-terminal, une perte de 160,0315 Da est mesurée entre l'ion parent à m/z 992,3740 et le premier fragment observé à m/z 832,3425. Dans les conditions de recherche (0,05 Da près), seul l'enchaînement Gly-Cys est retenu. Cette proposition est tout à fait acceptable, d'une part en raison de l'écart de masse observé entre expérience et théorie (5,3 ppm) et d'autre part parce que l'on attend deux cystéines dans la séquence complète.

Maintenant que la composition des deux extrémités de la toxine a été déterminée, une recherche est effectuée sur les différences de masses des ions correspondant plus au centre du peptide. Comme le montre le *Tableau 4*, aucun résultat réellement pertinent n'est relevé pour les différences de masse de 243,1375 et 185,0325 Da.

Deux paramètres peuvent expliquer cela. Tout d'abord, il est possible que certaines modifications post-traductionnelles atypiques n'aient pas été prises en compte dans nos recherches. Ceci expliquerait qu'aucune composition en acides aminés n'est déterminée pour

la partie « centrale » du peptide. L'autre possibilité est que l'ion à m/z 589,2066 n'appartienne pas à la même série d'ions que ceux étudiés jusqu'alors. Pour trouver la séquence manquante du peptide, il faudrait alors étudier la différence de masse entre les ions à m/z 832,3425 et à m/z 404,1747 soit un écart de 428,1700 Da. Cette masse étant plus élevée, la recherche à 0,05 Da près donne près de 90 solutions. Pour clarifier la présentation des résultats, seules les possibilités trouvées à 20 ppm près (à la masse 428,1700, cela correspond à 0,009 Da d'erreur acceptée) seront détaillées dans le *Tableau 5*.

Masse mesurée	Extrémité	Masse recherchée	Possibilités	Masses théoriques	Erreurs (ppm)
427,1700	--	428,1700	EWO	428,1696	0,9
			APEM/SPMO/AMOO	428,1729	6,8
			GNQE/ GDQQ/ ANNE/ ANDQ/		
			GGGQE/ GGANE/ GGADQ/		
			GAAND/ GGGGAE /	428,1656	10,3
			GGGAAD/ SNNO/ GGSNO/		
			GGGGSO		

Tableau 5 : Propositions de séquences pour la masse de 427,1700 Da - Seuls les résultats à moins de 20 ppm d'erreur sont représentés ici - O désigne une hydroxyproline

Même avec cette tolérance très basse de nombreuses possibilités sont trouvées. La meilleure d'entre elles est l'enchaînement Glu-Trp-Hyp mesuré à moins de 1 ppm près (Hyp = hydroxyproline). Bien que cette proposition soit intéressante, il est difficile de conclure définitivement sur la séquence en se basant sur cette seule observation. En effet, trois *tags* sont trouvés avec des erreurs de 7 ppm environ et 13 autres avec des erreurs de 10 ppm, ce qui constitue des précisions acceptables pour ce type d'expérience.

Néanmoins, il apparaît qu'à ce stade, les informations obtenues indiquent que la séquence **la plus probable** est : Gly+Cys/Glu+Trp+Hyp/Pro/Cys+Trp* (* = amidation C-terminale). Cette séquence est intéressante puisque la présence des deux prolines et d'un acide glutamique au sein de la séquence permettrait d'expliquer, comme dans le cas des toxines de *Conus virgo*, l'absence de certaines fragmentations. L'enchaînement des acides aminés ne peut cependant pas être déterminé plus précisément.

Pour tenter d'avancer dans la caractérisation, nos hypothèses de séquence ont été comparées aux séquences déjà décrites des différents contryphans et conopressines connues (*Figure 22*). Trois conopressines sont répertoriées. Il s'agit des conopressines-G, -S et -T provenant respectivement des venins de *Conus tulipa*, *geographus* et *striatus*¹⁶. Ce sont des peptides composés de 9 acides aminés contenant un pont disulfure et une amidation C-terminale. Ils

présentent quatre acides aminés conservés : les deux cystéines en position +1 et +6 ainsi qu'une isoleucine en +3 et une asparagine en +5. A part la cystéine en +1 qui pourrait correspondre, aucune homologie évidente n'est relevée entre ces trois séquences et celle de la toxine étudiée. Ceci nous permet d'écarter assez rapidement la famille des conopressines et de se focaliser sur celle des contryphans.

Cône	Famille	Nom	Séquence									Masse (Da)
Geographus	Conopressine	Conopressin-G	C	F	I	R	N	C	P	K	G*	1033,4950
Striatus	Conopressine	Conopressin-S	C	I	I	R	N	C	P	R	G*	1027,5168
Tulipa	Conopressine	Conopressin-T	C	Y	I	Q	N	C	L	R	V*	1107,5318
Radiatus	Contryphan	Contryphan-R	G	--	C	O	<u>W</u>	E	P	W	C*	989,3524
Radiatus	Contryphan	Bromo-contryphan	G	--	C	O	<u>W</u>	E	P	<u>W</u>	C*	1067,2629
Textile	Contryphan	Contryphan-R/Tx	G	--	C	O	<u>W</u>	E	P	W	C*	989,3524
Textile	Contryphan	Contryphan-Tx	G	--	C	O	<u>W</u>	Q	P	Y	C*	965,3524
Textile	Contryphan	Leu-Contryphan-Tx	--	C	V	<u>L</u>	Y	P	W	C*		879,3772
Purpurascens	Contryphan	Contryphan-P	G	--	C	O	<u>W</u>	D	P	W	C	976,3208
Purpurascens	Contryphan	Leu-Contryphan-P	G	--	C	V	<u>L</u>	L	P	W	C	887,4034
Stercusemuscarum	Contryphan	Contryphan-Sm	G	--	C	O	<u>W</u>	Q	P	W	C*	988,3684
Ventricosus	Contryphan	Contryphan-Vn	G	D	C	P	<u>W</u>	K	P	W	C*	1087,4368
Ermineus	?	?	[G+C]		[O+W+E]		P	[W+C]*				989,3490

Figure 22: Séquences et masses calculées des différents contryphans et conopressines connus - O désigne une hydroxyproline, W un bromocontryphan et * une amidation C-terminale – Les acides aminés de configuration D sont soulignés.

De nombreux contryphans sont d'ores et déjà connus^{17,18}. Il s'agit de petits peptides de taille comparable aux conopressines (9 ou 10 acides aminés) qui possèdent également un unique pont disulfure. Malgré cette petite taille, certains comportent un nombre important de modifications post-traductionnelles. C'est le cas notamment du bromo-contryphan¹⁹ de *Conus radiatus* qui présente une amidation C-terminale, un pont disulfure, une hydroxyproline et un bromotryptophane pour seulement neuf acides aminés ! Une autre particularité des ces peptides est que le tryptophane en position +5 n'est pas de configuration L comme les acides aminés naturels mais D^{20,21}. Bien que l'activité biologique des contryphans ne soit pas réellement bien définie, ces peptides ont déjà été utilisés pour mimer l'activité biologique d'autres toxines. En effet, la forte structuration apportée par le motif CPXXPXC permet d'obtenir des peptides capables de bloquer les canaux calciques de type N en mimant l'activité de l' ω -conotoxine GVIA²².

Par comparaison de nos hypothèses avec les séquences de contryphans connues, il apparaît que le contryphan R et le R/Tx, de structure identique, ont une masse identique (à 3,4 ppm près) de celle qui a été mesurée pour *Conus ermineus*. Leur séquence possède exactement les mêmes acides aminés que celle que nous proposons. On peut donc supposer que le

contryphan analysé dans *Conus ermineus* est le même que celui qui a déjà été décrit dans les venins de *Conus radiatus* et *Conus textile*. Le fait de trouver une même toxine au sein de deux venins d'espèces différentes est peu courant²³. Néanmoins, ce contryphan particulier a été caractérisé dans trois venins différents. Bien que son activité biologique ne soit pas déterminée, sa présence au sein des trois venins indique qu'il joue probablement un rôle important.

Après analyse plus précise du spectre de fragmentation (*Figure 23*), il apparaît que les deux ions principaux correspondent aux ions y_3 et y_6 tandis que l'ion à m/z 589,2066 correspond à l'ion b_6 (1,5 ppm). Trois ions internes, confirmant la séquence totale, sont retrouvés avec de bonnes précisions sur les mesures de masses (<10ppm).

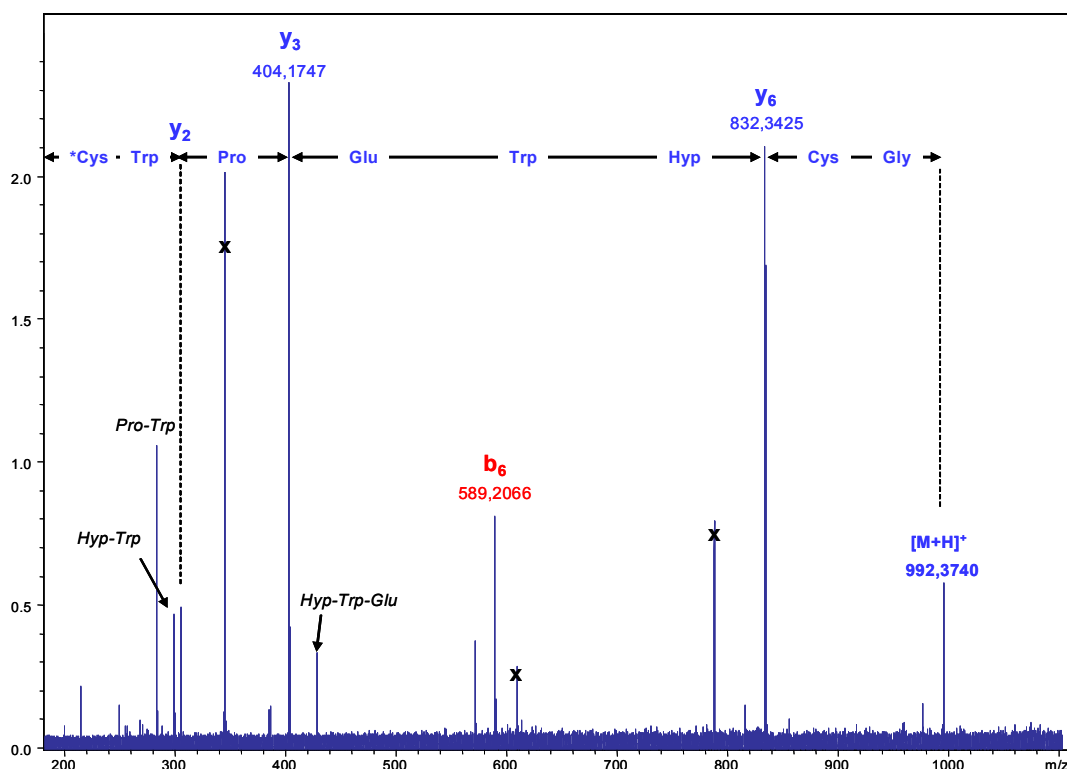


Figure 23: Spectre SORI-CAD annoté du contryphan Er₁K₉₈₉ – En italique sont indiqués les différents fragments internes informatifs



Lorsque l'on détaille maintenant les fragmentations obtenues, il apparaît clairement que, excepté pour la formation de l'ion y_2 , seules les liaisons *N*-terminales des prolines ont été clivées. Si l'on s'intéresse de plus près aux trois fragments internes, ces derniers incluent

chacun une fragmentation *N*-terminale des prolines, voire même deux dans le cas de l'ion à m/z 429,1711. Pour la formation de l'ion y_3 , on retrouve le phénomène lié à la présence du groupement Glu-Pro supposé dès le début de l'étude⁸.

Bien que peu d'ions fragments soient observés, le spectre de fragmentation est cohérent avec la séquence du contryphan même si aucune information ne nous permet de savoir si le tryptophane est bien de conformation D comme dans les autres contryphans. La présence de deux prolines dans sa séquence induit les fragmentations atypiques qui ont été expliquées.

C.2. Séquençage *de novo* dans la fraction D

C.2.a. Approche par FT-ICR

La seconde toxine potentielle, retrouvée dans la fraction D, est un peptide de masse 1772,7084 Da. Après réduction, sa masse augmente de quatre Dalton indiquant la présence de deux ponts disulfures dans sa structure (*Figure 24*). En suivant la même notation que celle utilisée pour le contryphan, cette toxine est baptisée Er₁D₁₇₇₂ pour cette étude.

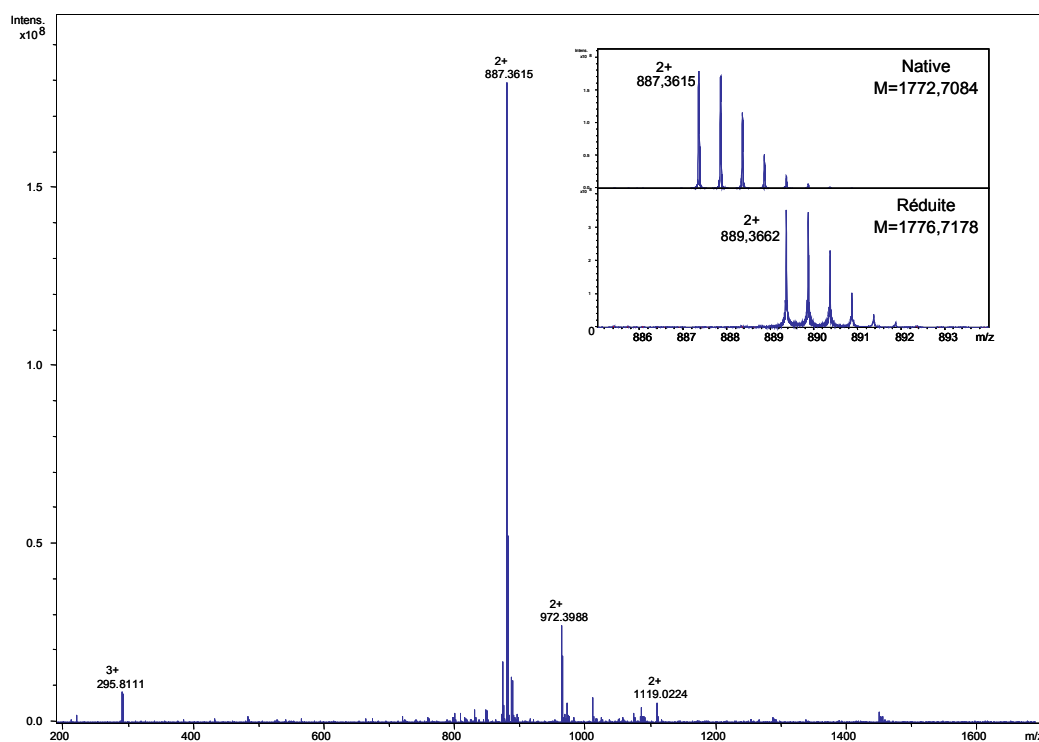


Figure 24: Spectre nanoESI-FT-ICR de la fraction Er₁D – En haut à droite, un zoom sur le signal du peptide d'intérêt obtenu avant et après réduction des ponts S-S

La fragmentation de $\text{Er}_1\text{D}_{1772}$ a été effectuée par collision avec du xénon en SORI-CAD après sélection de l'ion doublement chargé à m/z 889,3662 (*Figure 25*) et le spectre obtenu contient de nombreux ions d'intensités variables. Les essais de fragmentations par ECD n'ont pas permis l'obtention d'information complémentaire.

Une recherche automatique de *tags* de séquence utilisant une tolérance de 0,01 Da permet de trouver deux enchaînements d'acides aminés : Asn-Lys-Val-Cys et Asn-Trp-Cys. Chacun de ces *tags* est composé des ions les plus intenses du spectre et contient une cystéine. De nouveau, le manque d'ions fragments informatifs ne permet pas de trouver la séquence complète directement.

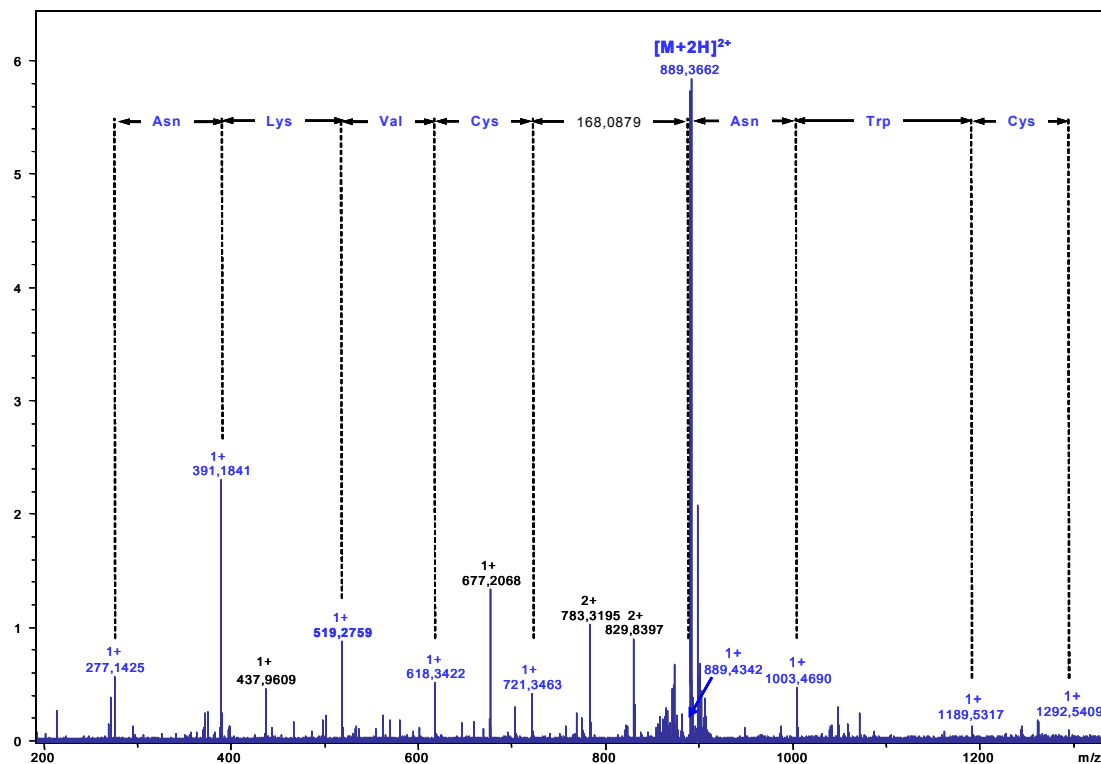


Figure 25: Spectre SORI-CAD (Xe) de l'ion (2+) à m/z 889,3662

Le séquençage de cette toxine a donc été abordé en suivant le même mode opératoire que les précédentes (*Tableau 6*). Il est important de rappeler que cette approche utilise l'hypothèse, non vérifiée *a priori*, que les séquences obtenues appartiennent à la même série d'ions (*y* ou *b* dans ce cas précis).

Masses mesurées	Extrémité	Masses recherchées	Possibilités	Masses théoriques	Erreurs (ppm)
168,0879	--	168,0879	AP	168,0899	11,9
			CR	259,1103	7,0
277,1423	C -ter (amidé)	259,1085	GTT/AST	259,1168	32,1
			GCV/QM/GAM	259,0991	36,2
485,1964	N -ter	485,1964	GCVO/APTCO/ GPCLD /GPVC	485,1944	4,1
			E/APVCD/SPPTC		
			CROO/PCER	485,2056	19,0

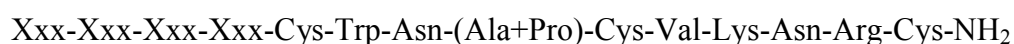
Tableau 6: Recherche de *tags* de séquence pour la fragmentation de l'ion m/z 889,36 (2+) par la précision en masse - O et Z désignent respectivement une hydroxyproline et un acide pyroglutamique. L correspond à Leu/Ile de manière indifférenciée

Pour la masse de 168,0879 Da séparant les deux tags trouvés par séquençage *de novo* automatique, seule une possibilité est trouvée. L'enchaînement Pro-Ala est déterminé avec une erreur de 11,9 ppm, ce qui peut paraître un peu élevé. Néanmoins, étant donné qu'aucune autre possibilité n'est trouvée par ailleurs, cette proposition peut être considérée comme valide. De plus, la présence d'une nouvelle proline, peut expliquer l'absence de fragments inter-résidus si l'enchaînement est Pro-Ala.

L'ion à m/z 277,1423 est caractéristique d'une extrémité terminale de la toxine. Comme précédemment, différentes possibilités pour les extrémités C-terminale (acide/amide) et N-terminale sont prises en compte. Pour les deux premières, la recherche de *tags* ne donne aucun résultat pertinent. En revanche, pour l'extrémité C-terminale amidée, l'enchaînement Cys-Arg est trouvé avec une précision acceptable (7 ppm) et apporte la troisième des quatre cystéines attendues. Les autres propositions dépassant 30 ppm d'erreur sont éliminées.

Enfin, il reste à déterminer l'extrémité N-terminale. En effet, en effectuant une recherche avec une tolérance de 0,01 Da pour déterminer la composition en acides aminés correspondant à une masse de 485,1964 Da, 51 réponses sont trouvées. En éliminant toutes celles qui ne possèdent pas de cystéines (condition essentielle à ce stade), il reste encore 8 possibilités (Tableau 6) et six de ces peptides sont retrouvés avec 4,1 ppm d'erreur. Il est donc impossible de conclure par cette approche sur la séquence de l'extrémité N-terminale.

Les informations obtenues nous conduisent donc à proposer la séquence suivante :



L'ordre des acides aminés composant l'extrémité C-terminale a pu être mis en évidence par la présence d'ions caractéristiques de fragmentations à m/z 399,2422 pour Lys-Asn-Arg et à m/z

769,4031 pour (Pro+Ala)-Cys-Val-Lys-Asn-Arg caractérisant la juxtaposition de la lysine et de l'asparagine C-terminales.

Pour déterminer la séquence totale du peptide, nous avons donc entrepris sa fragmentation avec un spectromètre de masse Q-TOF. Contrairement aux instruments FT-ICR qui permettent difficilement de détecter des ions de faible m/z , les spectromètres de masse Q-TOF conduisent à des spectres de fragmentation très riches, même dans les bas m/z , avec notamment la présence des ions immoniums. Par ailleurs, l'efficacité de la fragmentation dans la cellule de collision précédant le TOF est meilleure que dans une cellule ICR ce qui conduit à des ions fragments plus nombreux. Un dernier avantage est que la durée totale d'une acquisition en Q-TOF (1s environ) est bien plus courte que dans un ICR (5s en raison du temps de pompage du gaz inerte), ce qui permet d'effectuer plus d'accumulations (et donc d'améliorer le rapport signal/bruit) en un temps identique.

C.2.b. Approche par Q-TOF

Le spectre obtenu (fragmentation de l'ion (2+) à m/z 889,36) est très informatif et la quasi-totalité des ions y ainsi que quelques ions b de faibles intensité (non indiqués sur le spectre) sont obtenus (*Figure 26*).

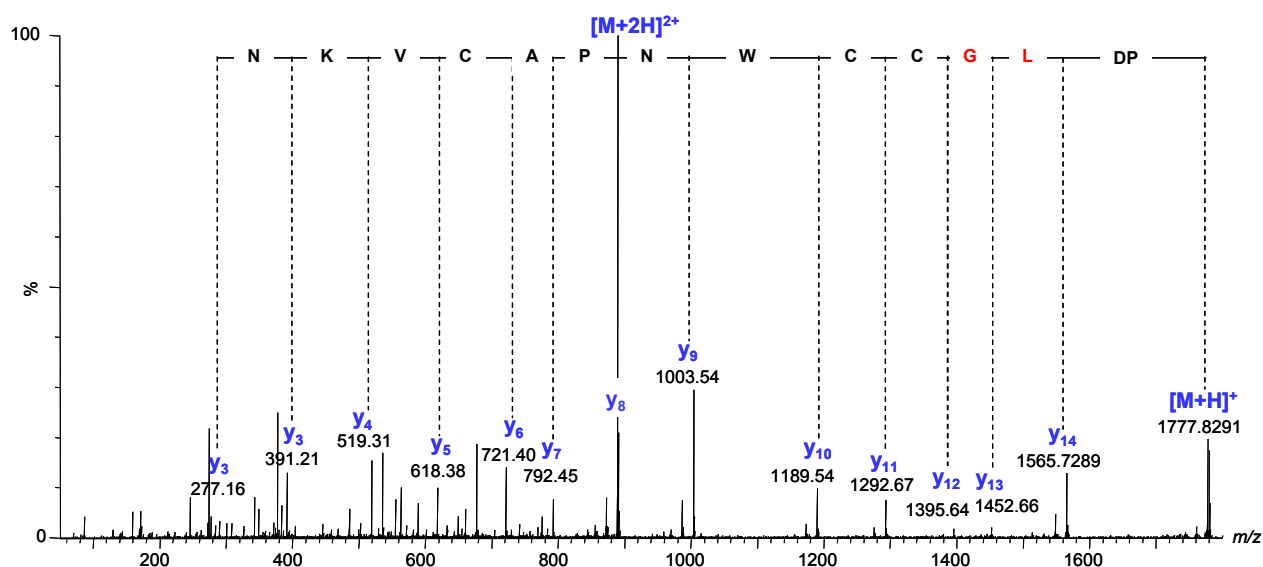


Figure 26: Spectre de fragmentation CID de Er₁D₁₇₇₂ sur le Q-TOF

Cette expérience apporte des informations qui valident parfaitement les observations effectuées sur les spectres FT-ICR. Tout d'abord, les ions qui ont permis initialement de déterminer des tags en acides aminés en FT-ICR sont bien retrouvés lors de cette expérience. Ensuite, un ion intense à m/z 792,45 permet, non seulement, de confirmer le tag Ala-Pro déterminé grâce à la précision sur la mesure de masses (*Tableau 6*) mais aussi de déterminer l'enchaînement de ce tag dans la séquence de la toxine. Enfin, la séquence *N*-terminale peut être déterminée ce qui constitue le point le plus important de cette étude. Une série d'ions y à m/z 1565,77, 1452,66, 1395,66 et 1292,67 permet de compléter par Cys-Gly-Leu/Ile la partie *N*-terminale du peptide. Ainsi, un seul des *tags* proposés dans le *Tableau 6* pour le *N*-terminal reste valide et permet de conclure quant à la séquence de la toxine :

Pro-Asp-Leu/Ile-Gly-Cys-Cys-Trp-Asn-Pro-Ala-Cys-Val-Lys-Asn-Arg-Cys-NH₂

Le spectre de fragmentation SORI-CAD est analysé de nouveau en prenant en compte la séquence précédente. Une série d'ions b , de faible intensité dans le spectre, et qui n'avait de ce fait pas été prise en compte lors de la première analyse, apparaît clairement (*Figure 27*).

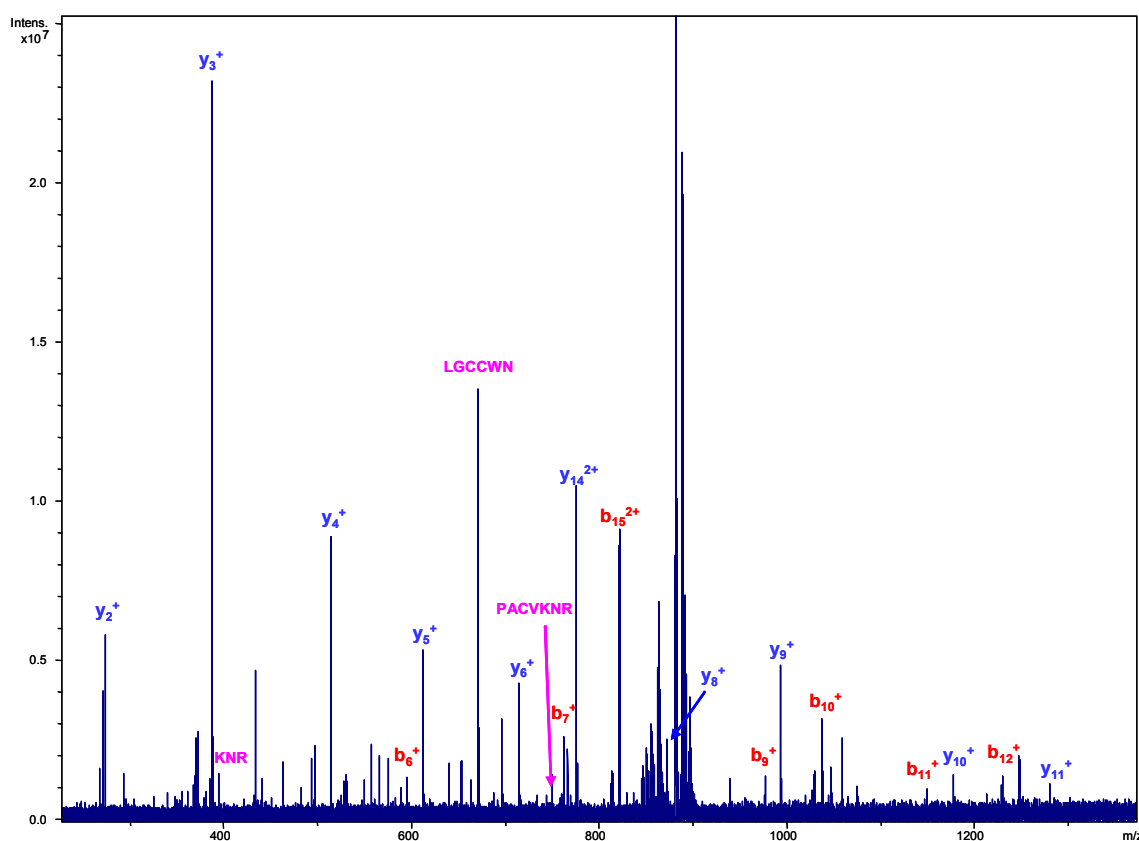
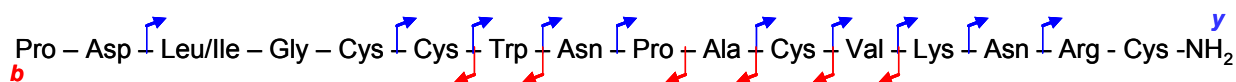


Figure 27: Spectre de fragmentation SORI-CAD de Er₁D₁₇₇₂ annoté



Afin de déterminer la famille de cette nouvelle toxine, la séquence est comparée à celles de conotoxines possédant deux ponts disulfures (*Tableau 7*).

Cône	Famille	Nom	Séquence																								
Marmoreus	χ	MrlA	N	G	V	C	C	G	Y	K	L	C	H	O	C												
Marmoreus	χ	MrlB	V	G	V	C	C	G	Y	K	L	C	H	O	C												
Textile	T-superfamille	Tx5.2				E	C	C	E	D	G	W	C	C	T	A	A	P	L	T	G	R					
Pupurascens	T-superfamille	P5a				G	C	C	P	K	Q	M	R	C	C	T	L	*									
Aulicus	T-superfamille	Au5a				F	C	C	P	F	I	R	Y	C	C	W											
Gloriamaris	T-superfamille	Gm5.2				V	C	C	R	P	V	Q	D	C	C	S	G	K									
Virgo	T-superfamille	ViVI A				p	C	C	I/L	T	I/L	E	P	C	C	R	I/L	*									
Virgo	T-superfamille	ViVI B				p	C	C	I/L	T	P	E	P	C	C	R	V	*									
Aulicus	α	AulA				G	C	C	S	Y	P	P	C	F	A	T	N	S	D	Y	C	*					
Bullatus	α	BulA				G	C	C	S	T	P	P	C	A	V	L	Y						C	*			
Conors	α	CnlA	G			R	C	C	H			P	A	C	G	K	Y	Y	S				C	*			
Episcopatus	α	Epl				G	C	C	S	D	P	R	C	N	M	N	N	P	D	Y*			C	*			
Imperialis	α	ImI				G	C	C	S	D	P	R	C	A	W	R							C	*			
Regius	α	RglA				G	C	C	S	D	P	R	C	R	Y	R							C	R	*		
Geographus	α	GI				E	C	C	N			P	A	C	G	R	H	Y	S					C	*		
Geographus	α	GIA				E	C	C	N			P	A	C	G	R	H	Y	S					C	G	K	*
Magus	α	MII				G	C	C	S	N	P	V	C	H	L	E	H	S	N	L			C	*			
Pupurascens	α	PIA	R			D	P	C	C	S	N	P	V	C	T	V	H	N	P	Q	I			C			
Ermineus	α ?	?	P	D	L	G	C	C	W	N	P	A	C	V	K	N	R							C	*		

Tableau 7: Comparaison de la séquence de la toxine Er₁D₁₇₇₂ avec celle de conotoxines appartenant à différentes familles – O désigne une hydroxyproline – Y* une tyrosine phosphorylée – l'astérisque finale * indique une amidation C-terminale

Il apparaît rapidement que l'agencement des ponts de la conotoxine que nous venons de séquencer dans le venin de *Conus ermineus* n'est pas compatible avec celui de la superfamille T et des χ -conotoxines qui présentent des motifs de type CC...CC ou CC...CHOC. En revanche, le motif présenté par notre nouvelle conotoxine apparaît bien plus semblable à celui des α -conotoxines. Dans cette famille, les deux premières cystéines sont vicinales, elles sont séparées de la troisième cystéine par trois ou quatre acides aminés et cette troisième cystéine est elle-même séparée de la dernière par trois à sept aminés. Notre conotoxine s'inscrit parfaitement dans cet agencement. L'une des particularités des α -conotoxines est de posséder une proline très conservée (*Tableau 7* en rouge). Cette proline est non seulement présente dans notre toxine mais est située entre une alanine et une asparagine formant un tripeptide NPA déjà observé plusieurs fois au sein des α -conotoxines (GI et GIA dans nos exemples).

De nombreuses α -conotoxines ont déjà été décrites mais cette nouvelle séquence apporte une diversité supplémentaire à la famille des α -conotoxines. L'extrémité N-terminale est la plus longue jamais décrite et l'introduction d'un tryptophane volumineux et hydrophobe au centre de la boucle de la toxine est très atypique. Pour évaluer l'impact biologique de ces deux observations la toxine a été synthétisée puis soumise à une recherche d'activité biologique.

C.2.c. Activité biologique de l' α -conotoxine Er₁D₁₇₇₂

La toxine séquencée est synthétisée puis purifiée par chromatographie à polarité de phase inverse. Comme nous l'avons déjà abordé, les α -conotoxines sont actives sur les récepteurs à l'acétylcholine. Plus précisément, les α -conotoxines qui possèdent le *tag* NPA (ou HPA) dans leur séquence présentent des affinités bien supérieures pour les AChR situés au niveau musculaire. De ce fait, l'activité de Er₁D₁₇₇₂ a été évaluée par des tests de liaisons sur ces récepteurs musculaires. Dans ce type d'expérience, la valeur de référence est la concentration de compétiteur froid pour laquelle 50% de la liaison spécifique de ligand marqué est déplacée. Elle est couramment appelée IC₅₀ pour concentration à 50% d'inhibition ; plus l'IC₅₀ est faible et plus le ligand aura une grande affinité pour le récepteur. Le modèle choisi pour ces expériences est celui des récepteurs de Torpille (raie électrique). Ces récepteurs, analogues aux récepteurs nicotiniques musculaires (AChR pouvant fixer acétylcholine et nicotine), sont mis au contact d'un ligand radioactif iodé hautement sélectif : l' α -bungarotoxine, extraite du venin du serpent *Bungarus multicinctus*. Les résultats de cette étude (effectuée par le Dr. Denis Servent, CEA Saclay) sont présentés *Figure 28*.

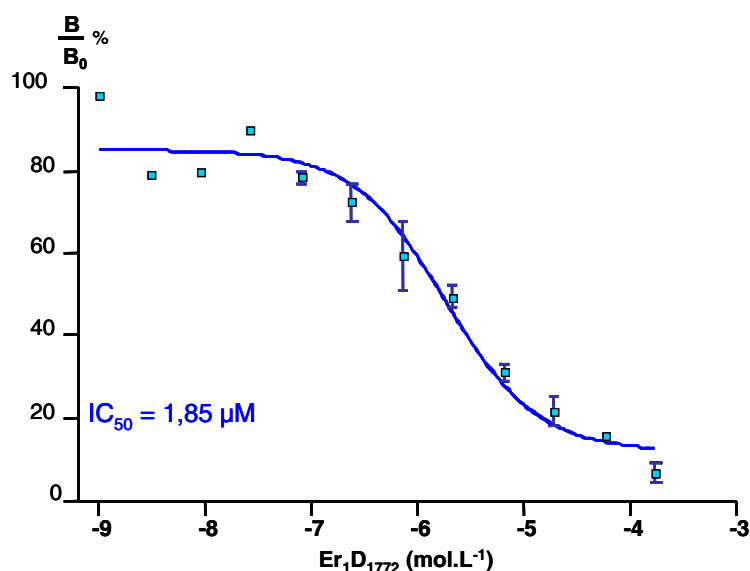


Figure 28: Affinité de Er₁D₁₇₇₂ pour les récepteurs à l'acétylcholine

Cette expérience montre que la conotoxine Er₁D₁₇₇₂ perturbe la liaison entre les AChR de torpille et l' α -bungarotoxine avec une affinité élevée (IC₅₀ est mesurée à 1,85 μ M). Néanmoins, il s'agit ici d'expériences préliminaires qui nécessitent d'être répétées plusieurs

fois et l'intérêt principal est de déterminer si cette nouvelle toxine est capable de discriminer les deux sites de fixation connus sur le récepteur musculaire comme le font notamment les conotoxines GI et MI²⁴. Des études complémentaires permettront de répondre à ces questions. Parallèlement des études en électrophysiologie ont été effectuées par le Dr Jordi Molgo (CNRS, Gif-sur-Yvette). Les résultats montrent que l'Er₁D₁₇₇₂ bloque les récepteurs nicotiques à l'acétylcholine musculaires ce qui se traduit par le blocage de la transmission neuromusculaire *in vitro*. Sur une proie l'effet de cette toxine serait une diminution puis un blocage de la contraction musculaire sans affecter la conduction nerveuse. Cette activité est tout à fait caractéristique des α -conotoxines (possédant NPA ou HPA) déjà connues. Bien que les expériences préliminaires n'aient pas démontré une activité originale de Er₁D₁₇₇₂, d'autres expériences restent à effectuer comme étudier sa discrimination au niveau des deux sites de fixation du récepteur musculaire.

C.3. Le venin de *Conus ermineus*, une richesse insoupçonnée...

Outre les deux toxines analysées dans ce travail, l'approche par chromatographie a permis de pointer plusieurs autres composés dont l'étude, non achevée, ne sera qu'abordée ici.

La conotoxine δ -EVIA (mise en évidence par le CEA très récemment) est actuellement le composé issu de *Conus ermineus* qui présente le plus grand intérêt. Après marquage fluorescent, sa sélectivité pour les canaux ioniques neuronaux a été exploitée pour en faire le premier marqueur sélectif de ces canaux. La recherche d'autres toxines de ce type constitue donc une grande motivation. Une toxine de 3525,4930 Da a été sélectionnée grâce aux analyses chromatographiques pour ses homologies avec la conotoxine δ -EVIA (M= 3285,4459 Da). En effet, non seulement ses trois ponts disulfures mais aussi sa masse voisine permettent de faire de cette toxine un très bon candidat. Après recherche dans les fractions de purification, la toxine est retrouvée dans la fraction F, et est nommée Er₁F₃₅₂₅ (Figure 29).

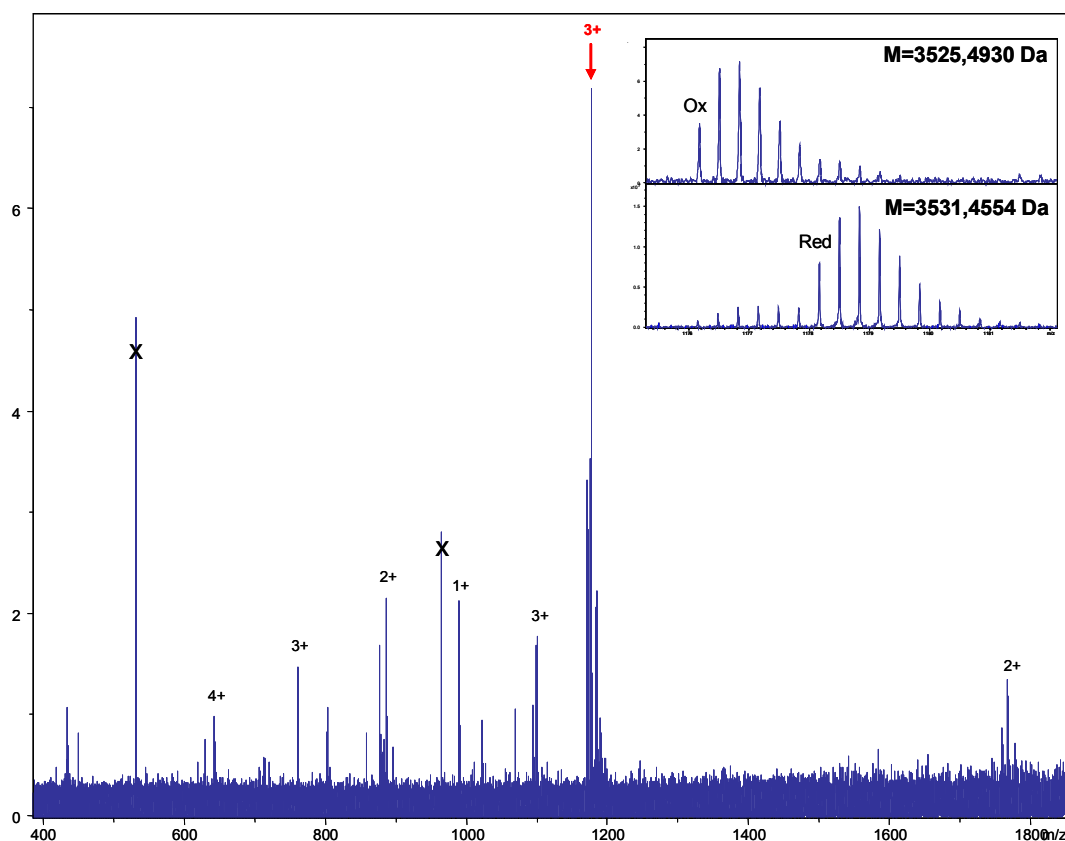


Figure 29: Spectre nanoESI-FT-ICR de la fraction Er1F – la toxine Er₁F₃₅₂₅ est indiquée d’une flèche rouge – L’encadré représente son massif isotopique avant et après réduction au TCEP, le décalage observé révèle la présence de trois ponts disulfures

Des expériences préliminaires de fragmentation par ECD ont été menées sur cette toxine. Le principal tag déterminé par cette expérience est un enchaînement Pro-Glu-Cys-Cys-Ser-Asn. Une recherche par BLAST nous montre que 8 des 12 meilleurs résultats concernent des toxines de cônes. Sur ces 8, 4 appartiennent à la superfamille O (δ , μ O ; κ et ω), 3 sont des α -conotoxines, et la dernière est une γ -conotoxine. Ces informations ne permettent pas de confirmer à quelle famille appartient précisément cette nouvelle conotoxine. Le tag de 5 acides aminés trouvé est trop petit pour être plus restrictif sur la recherche dans les bases de données et de nouvelles expériences sur une quantité plus importante de fraction devront être menées pour continuer sur cette piste.

Une autre piste ouverte grâce à notre criblage du venin par chromatographie est celle des toxines glycosylées. Les toxines concernées sont trouvées dans la fraction ER₁I (Figure 30). La masse mesurée pour la plus petite d’entre elles est de 8350,1493 Da.

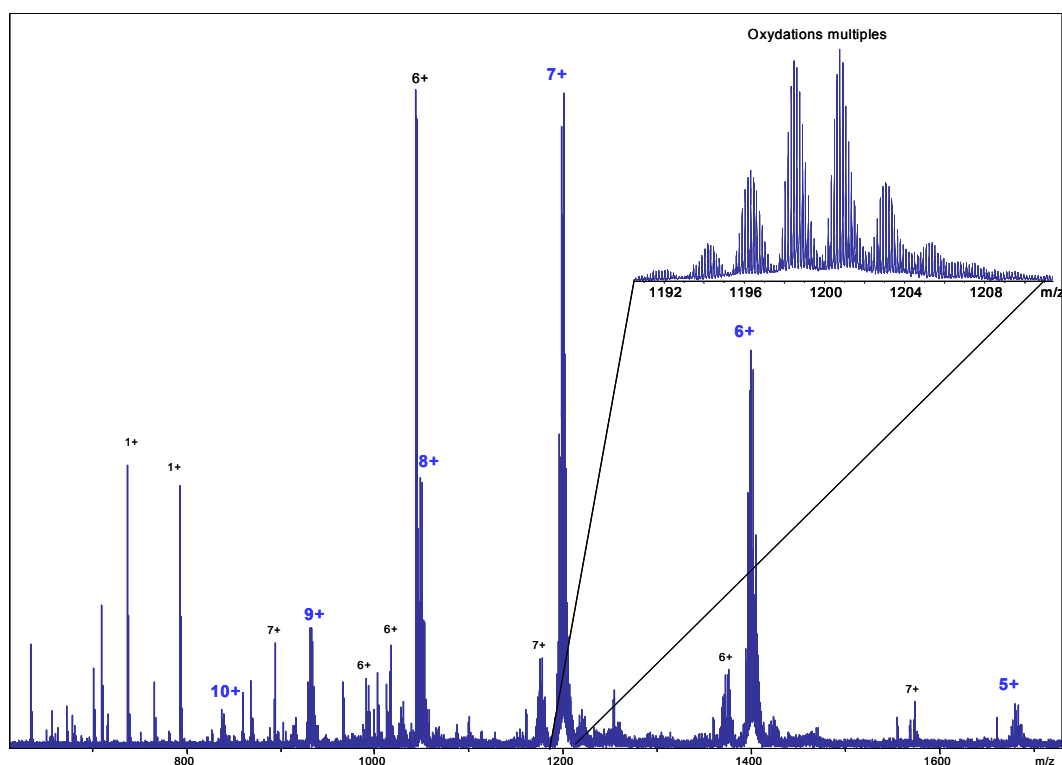


Figure 30: Spectre nanoESI-FT-ICR de la fraction Er1I – les massifs isotopiques des $[M+7H]^{7+}$ des toxines glycosylées montrent la présence de nombreuses oxydations

L'analyse du spectre montre que cette toxine possède de nombreuses formes oxydées puisque plusieurs massifs isotopiques consécutifs sont décalés de 16 Da. Ceci est probablement dû à la présence importante de méthionines ou de tryptophane, acides aminés susceptibles de s'oxyder.

La réduction des ponts disulfures indique un décalage de 12 Da, caractéristique de la présence de 6 ponts disulfures dans la séquence de cette toxine. Même si des toxines d'araignées²⁵ et de serpents²⁶ possédant 6 ponts disulfures ont été décrites dans des gammes de masses tout à fait similaires, aucune famille de conotoxines n'en possède autant. Les σ -conotoxines, avec 5 ponts, sont celles qui en présentent le plus.

Une expérience de dissociation en source (ISD) a été menée sur cet échantillon en ayant comme but initial d'obtenir un maximum d'informations sur la séquence (Figure 31). Le spectre obtenu est très complexe en raison de la fragmentation de l'ensemble des molécules présentes dans la fraction I. Bien qu'aucun élément de séquence ne puisse être déterminé, on peut noter la présence d'intenses fragmentations dans la gamme $[m/z\ 1500 - m/z\ 1700]$.

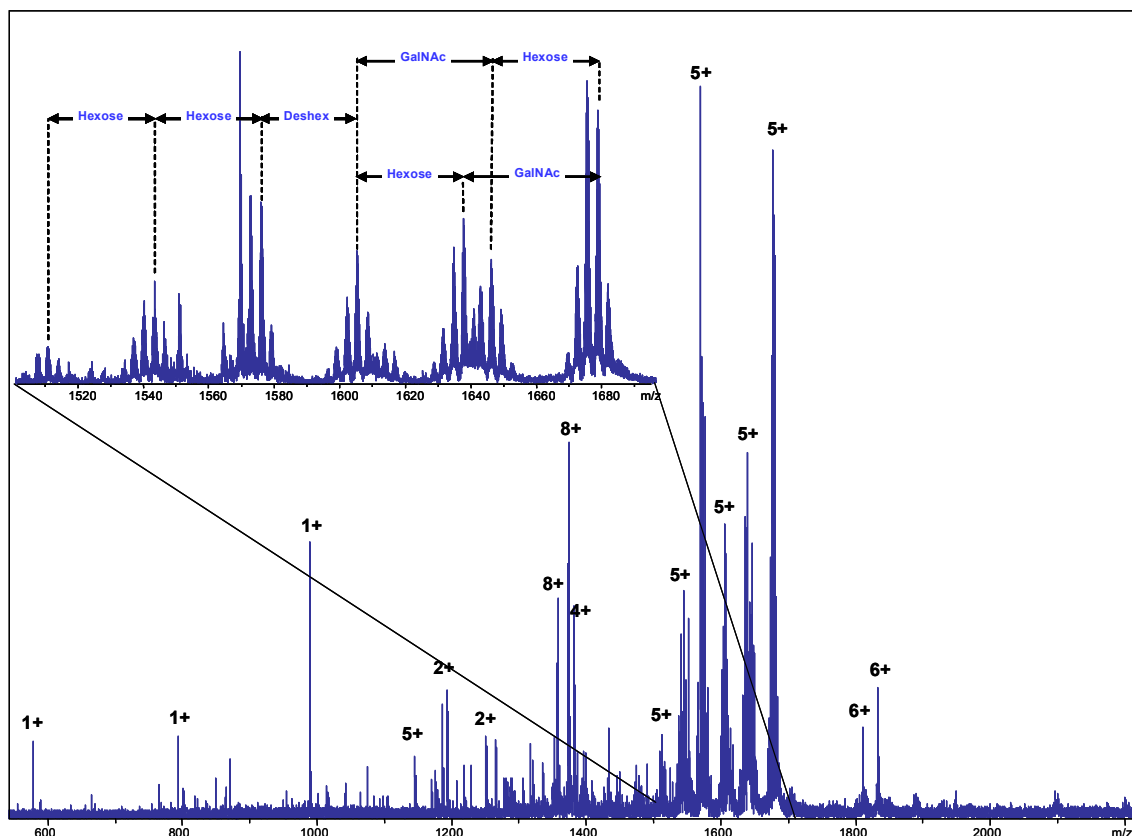


Figure 31: Spectre ISD (200V, 100 scans) de la fraction ErII indiquant la présence de glycanes sur la famille de composés vers 8350 Da – Un zoom est effectué sur la gamme m/z [1500-1700]

Ces fragments correspondent à la perte de différents motifs de glycanes à partir des peptides observés majoritairement. La présence de toxines glycosylées (O- et N-glycosylées) est très rare mais a déjà été décrite. En effet, la présence de sucres a déjà été identifiée dans des séquences de toxines issues de venins de cônes²⁷, d'araignée²⁵ et de scorpion²⁸. La détermination précise de l'état de glycosylation d'une toxine de 9 kDa représente un travail important nécessitant une quantité de matériel non négligeable. Plusieurs approches peuvent être envisagées : « bottom-up » avec digestion de la protéine ou « top-down » sur protéine entière, impliquant l'utilisation d'ECD. Ce travail n'a pu être réalisé dans le cadre de cette thèse mais les deux approches complémentaires sont envisagées.

D. Conclusion

L'analyse du venin brut de *Conus ermineus* a été réalisée par plusieurs approches. La première en infusion directe par nanoESI n'a pu conduire à des résultats satisfaisants en raison de la complexité du venin étudié. La seconde approche repose sur l'utilisation d'un

couplage en ligne nanoLC(RP)-FT-ICR. Elle a permis d'une part de mesurer 370 masses moléculaires différentes et d'autre part de déterminer le nombre de ponts disulfures pour plus d'un tiers de ces composés. Comme attendu pour les conotoxines, on observe une majorité de masses entre 1000 et 5000 Da. Néanmoins, l'impossibilité de pouvoir réaliser des expériences de MS/MS en ligne sur notre instrument nous a conduit à découpler la séparation de l'analyse par spectrométrie de masse.

Les résultats obtenus sur les fractions sont tout à fait intéressants puisque deux nouvelles conotoxines ont été séquencées. Pour la première (contryphan, 1 pont disulfure, M=989,3490 Da), la précision en masse de la FT-ICR a permis de démontrer la que cette toxine a une forte probabilité de présenter la même séquence que l'un des contryphans déjà décrit alors que pour la seconde (α -conotoxine, 2 ponts disulfures, M=1772,7084 Da), les informations incomplètes obtenues grâce au spectre SORI-CAD nous a conduit a refaire les expériences en MS/MS sur un autre spectromètre de masse. Finalement, les résultats obtenus sur un Q-TOF ont été suffisants pour compléter la séquence de cette toxine. L'activité biologique de cette nouvelle α -conotoxine est en cours d'évaluation mais des expériences préliminaires montrent qu'elle possède un profil classique pour cette famille de conotoxines.

IV. Conclusion générale

L'ensemble des résultats obtenus pour *Conus virgo* et *Conus ermineus* montre à la fois les intérêts et les limitations de l'utilisation de la FT-ICR pour l'analyse d'un venin brut.

Pour établir la cartographie massique d'un venin, nous avons réalisé deux types d'expériences différentes : infusion directe du venin et couplage en ligne nanoLC-FT-ICR. Nos résultats montrent que si la première approche est envisageable pour un venin dont la complexité est moyenne, le couplage devient nécessaire lorsque cette complexité augmente. Les expériences de couplage montrent en effet qu'un nombre de composés plus important qu'en infusion directe peut être détecté. Malgré tout, certains composés sont perdus et les conditions expérimentales de la séparation peuvent certainement être améliorées pour éviter ce problème. L'apport de la FT-ICR pour établir une cartographie concerne la précision sur la mesure de masse et la résolution, que ce soit en couplage ou pas. En revanche, la faible gamme dynamique du FT-ICR est un point négatif pour l'étude de ces mélanges complexes.

En ce qui concerne le séquençage *de novo*, nous montrons qu'il est possible de séquencer de petits peptides (majoritaires dans le mélange) sans séparation préalable. Là encore, la précision de la FT-ICR est primordiale, principalement lorsque les spectres de fragmentation présentent peu de fragments. Nos résultats obtenus sur le Q-TOF pour le séquençage d'une toxine montrent que finalement, la méthodologie la mieux adaptée à ce type de problématique serait un instrument FT-ICR hybride de type Q-FT-ICR par exemple dans la mesure où cet instrument allie les qualités de la fragmentation d'un quadrupole à la précision et la résolution en masse d'un FT-ICR.

Sur le plan biologique, nous avons déterminé la séquence de deux nouvelles conotoxines originales appartenant à la superfamille T dans le venin de *Conus virgo* et mis en évidence la présence d'un contryphan et d'une nouvelle α -conotoxine, dans le venin de *Conus ermineus*. Les perspectives les plus intéressantes concernent sans doute la détection dans ce dernier venin de composés glycosylés de haut poids moléculaire. En effet, peu de peptides de ce type ont déjà été décrits dans les venins de cône.

D'un point de vue méthodologique le couplage nanoLC-Q-TOF disponible depuis peu de temps au laboratoire devrait permettre d'augmenter le débit de ces analyses et d'apporter rapidement des éléments de séquence pour l'ensemble des constituants d'un venin brut, et pas seulement des majoritaires.

Références Bibliographiques

- ¹ Corpuz G.P., Jacobsen R.B., Jimenez E.C., Watkins M., Walker C., Colledge C., Garrett J.E., McDougal O., Li W., Gray W.R., Hillyard D., Rivier J., McIntosh J.M., Cruz L.J., Olivera B.M., Definition of the M-conotoxin superfamily: characterization of novel peptides from molluscivorous *Conus* venoms, **2005**, *Biochem.*, 44, 8176-8186.
- ² Kauferstein S., Huys I., Lamthanh H., Stöcklin R., Sotto F., Menez A., Tytgat J., Mebs D., A novel conotoxin unhibiting vertebrate voltage-sensitive potassium channels, **2003**, *Toxicon*, 42, 43-52.
- ³ Kauferstein S., Melaun C., Mebs D., Direct cDNA cloning of novel conopeptide precursors of the O-superfamily, 2005, *Peptides*, 26, 361-367.
- ⁴ Arias H.R., Blanton M.P., α -Conotoxins, **2000**, *Int. J. Biochem. Cell Biol.*, 32, 1017-1028.
- ⁵ Walker C.S., Steel D., Jacobsen R.B., Lirazan M.B., Cruz L.J., Hooper D., Shetty R., DelaCruz R.C., Nielsen J.S., Zhou L.M., Bandyopadhyay P., Craig A.G., Olivera B.M., The T-superfamily of Conotoxins, **1999**, *J. Biol. Chem.*, 274, 30664-30671.
- ⁶ Dwoskin L.P., Crooks P.A., Competitive neuronal nicotinic receptor antagonists: a new direction for drug discovery, **2001**, *J. Pharmacol. Exp. Ther.*, 298, 395-402.
- ⁷ Sharpe I.A., Loughnan M.L., Thomas L., Adams D.A., Atkins A., Palant E., Craik D.J., Alewood P.F., Lewis R.J., Two new classes of conopeptides inhibit the α -1-adrenoreceptor and noradrenaline transporter, **2001**, *Nat. Neurosci.*, 4, 902-907.
- ⁸ Wysocki V.H., Tsaprailis G., Smith L.L., Brei LA.A, Mobile and localized protons : a framework for understanding dissociation, **2000**, *J. Mass. Spectrom.*, 35, 1399-1406.
- ⁹ Brei L.A., Tabb D.L., Yates III, J.R., Wysocki V.H., Cleavage N-terminal to proline : analysis of a database of peptide tandem mass spectra, **2003**, *Anal. Chem.*, 75, 1963-1971.
- ¹⁰ Wilm M., Mann M., Analytical properties of the nanoelectrospray source, **1996**, *Anal. Chem.*, 68, 1-9.
- ¹¹ Juraschek R., Dulcks T., Karas M., Nanoelectrospray – More than just a minimized-flow electrospray ion source, **1999**, *J. Am. Mass Spectrom.*, 10, 300-308.
- ¹² Taggi A.E., Meinwald J., Schroeder F.C., A new approach to natural products discovery discovery exemplified by the identification of sulphated nucleosides in spider venom, **2004**, *J. Am. Chem. Soc.*, 126, 10364-10396.
- ¹³ Barbier J., Lamthanh H., Le Gall F., Favreau P., Benoit E., Chen H., Gilles N., Ilan N., Heinemann S.H., Gordon D., Menez A., Molgo J., A δ -conotoxin from *Conus ermineus* venom inhibits inactivation in vertebrate neuronal Na⁺ channels but not in skeletal and cardiac muscles, **2004**, *J. Biol. Chem.*, 279, 4680-4685.
- ¹⁴ Volpon L., Lamthanh H., Barbier J., Gilles N., Molgo J., Menez A., Lancelin J.M., NMR solution structure of d-conotoxin EVIA from *Conus ermineus* that selectively acts on vertebrate neuronal NA⁺ channels, **2004**, *J. Biol. Chem.*, 279, 21356-21366.
- ¹⁵ Jacubowski J.A., Sweedler J.V., Sequencing and mass profiling highly modified conotoxins using global reduction alkylation followed by mass spectrometry, **2004**, *Anal. Chem.*, 76, 6541-6547.

- ¹⁶ Cruz L.J., De Santos V., Zafaralla G.C., Ramilo C.A., Zeikus R., Gray W.R., Olivera B.M., Invertebrate vasopressin/oxytocin homologs, Characterization of peptides from *Conus geographus* et *Conus striatus* venoms, **1987**, *J. Biol. Chem.*, 262, 15821-15824.
- ¹⁷ Pallaghy P.K., He W., Jimenez E.C., Olivera B.M., Norton R.S., Structures of the contryphan family of cyclic peptides. Role of electrostatic interaction in cis-trans isomerism, **2000**, *Biochem.*, 39, 12845-12852.
- ¹⁸ Jimenez E.C., Watkins M., Juszczak L.J., Cruz L.J., Olivera B.M., Contryphans from *Conus textile* venom ducts, **2001**, *Toxicon*, 39, 803-808.
- ¹⁹ Jimenez E.C., Craig A.G., Watkins M., Hillyard D.R., Gray W.R., Gulyas J., Rivier J.E., Cruz L.J., Olivera B.M., Bomocontryphan: Posttranslational bromination of tryptophan, **1997**, *Biochem.*, 36, 989-994.
- ²⁰ Jimenez E.C., Olivera B.M., Gray W.R., Cruz L.J., Contryphan is a D-tryptophan containing *Conus* peptides, **1996**, *J. Biol. Chem.*, 271, 28002-28005.
- ²¹ Pallaghy P.K., Melnikova A.P., Jimenez E.C., Olivera B.M., Norton R.S., Solution structure of Contryphan-R, a naturally occurring disulfide-bridged octapeptide containing D-tryptophan: Comparison with protein loops, **1999**, *Biochemistry*, 38, 11553-11559.
- ²² Pallaghy P.K., Norton R.S., The cyclic contryphan motif CPxXPXC, a robust scaffold potentially useful as an ω -conotoxin mimic, **2000**, *Biopolymers*, 54, 173-179.
- ²³ Escoubas P., Celerier M.L., Nakajima T., High-performance liquid chromatography matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry peptide fingerprinting of tarantula venoms in the genus *Brachypelma*: chemotaxonomic and biochemical applications, **1997**, *Rapid Comm. Mass Spectrom.*, 11, 1891-1899.
- ²⁴ Hann R.M., Pagan O.R., Eterovic V.A., The alpha-conotoxins GI and MI distinguish between the nicotinic acetylcholine receptor agonist sites while SI does not, **1994**, *Biochem*, 33, 14058-14063.
- ²⁵ Sutton K.G., Siok C., Stea A., Zamponi G.W., Heck S.D., Volkmann R.A., Ahljanian, Snutch T.P., Inhibition of neuronal calcium channels by a novel peptide spider toxin, DW13.3, 1998, *Mol. Pharmacol.*, 54, 407-418.
- ²⁶ Huang T.F., What have snakes taught us about integrins?, **1998**, *Cell. Mol. Life Sci.*, 54, 527-540.
- ²⁷ Craig A.G., Zafaralla G., Cruz L.J., Santos A.D., Hillyard D.R., Dykert J., Rivier J.E., Gray W.R., Imperial J., Delacruz R.G., Sporning A., Terlau H., West P.J., Yoshikami D., Olivera B.M., An O-glycosylated neuroexcitatory *Conus* peptide, **1998**, *Biochem.*, 37, 16019-16025.
- ²⁸ Hassani O., Loew D., Van Dorsselaer A., Papandreou M.J., Sorokine O., Rochat H., Sampieri F., Mansuelle P., Aah VI, a novel, N-glycosylated anti-insect toxin from *Androctonus australis hector* scorpion venom : isolation, characterization, and glycan structure determination, **1999**, *Fed. Eur. Biochem. Soc. Lett.*, 443, 175-180.

Chapitre III

Approches « bottom-up » et « top-down » par
spectrométrie de masse FT-ICR pour la
caractérisation d'une toxine de serpent spécifique
du récepteur α_{1A} -adrénergique

I. Introduction

Comme nous l'avons vu au chapitre précédent, l'analyse de la structure primaire de petites toxines peptidiques peut se faire directement par spectrométrie de masse en tandem, sans étape de digestion préalable. Néanmoins, une difficulté apparaît lorsque que les toxines dépassent 3-4 kDa, qui est la limite pour que la fragmentation par collision de basse énergie avec un gaz neutre soit efficace et conduise à des spectres de fragmentation informatifs. Ceci est particulièrement vrai lorsque les espèces fragmentées sont monochargées. Les ions multichargés posent également problème car les spectres de fragmentation sont souvent difficiles à interpréter sur les instruments à basse résolution en raison de la présence d'ions multichargés dans les spectres.

Deux approches basées sur différentes techniques de spectrométrie de masse peuvent être envisagées pour les toxines de haut poids moléculaire. La première, « bottom-up » consiste à digérer la toxine par une enzyme (la trypsine en général) afin d'obtenir des peptides qui peuvent eux-mêmes être séquencés. La seconde, « top-down » consiste à fragmenter directement la protéine en phase gazeuse *via* des techniques de type capture d'électron dissociative (ECD).

Dans ce chapitre, le potentiel des deux méthodes sera évalué pour l'étude d'une toxine de 7,3 kDa extraite et purifiée du venin du mamba vert, *Dendroaspis angusticeps*. L'intérêt de l'approche « top-down » utilisée pour contrôler les différentes étapes de la transformation de cette toxine en une sonde fluorescente spécifique des récepteurs α_{1A} -adrénergiques sera dégagé. Toutes les expériences qui ont permis la détermination de l'activité biologique et de la sélectivité de la toxine ont été effectuées par le Dr Nicolas Gilles du Département d'Ingénierie et d'Etude des Protéines (DIEP) du CEA à Saclay.

II. Les mambas

A. Généralités

Les mambas sont des serpents appartenant à la famille des élapidés (*Elapidae*). Cette famille comprend également le serpent corail et les cobras. Quatre espèces de mambas sont répertoriées, toutes en Afrique tropicale et sub-saharienne : le mamba noir (*Dendroaspis*

polylepsis), le mamba de Jameson (*Dendroaspis jamesonii*), le mamba vert de l'ouest (*Dendroaspis viridis*) et le mamba vert de l'est (*Dendroaspis angusticeps*) (Figure 1).

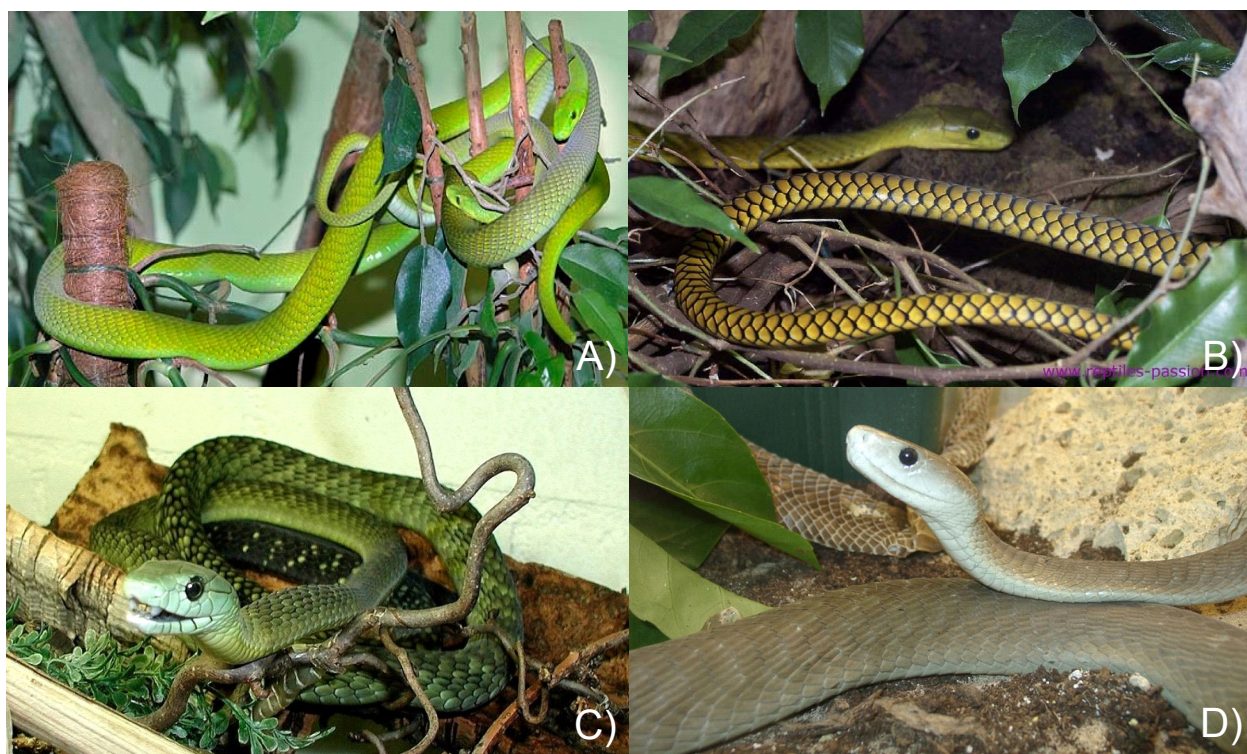


Figure 1: Les quatre mambas – (A) *Dendroaspis angusticeps* (B) *Dendroaspis viridis* (C) *Dendroaspis jamesonii* et (D) *Dendroaspis polylepsis*

Le mamba noir vit principalement sur le sol bien qu'il puisse se réfugier dans les arbres. Mesurant jusqu'à plus de quatre mètres, il constitue le deuxième serpent le plus long connu après le cobra royal *Ophiophagus hannah*. Les autres mambas sont environ deux fois plus petits, bien que *Dendroaspis angusticeps* dépasse fréquemment trois mètres de longueur. Compte tenu qu'ils possèdent un venin très puissant et qu'ils sont très rapides dans leurs mouvements (jusqu'à 16 km.h⁻¹ pour le mamba noir), les mambas sont catalogués comme des animaux extrêmement dangereux. La mortalité observée après une envenimation par le mamba noir est proche de 100% même lorsque les victimes reçoivent un traitement médical rapide.

B. Composition des venins des mambas

Le venin de ces serpents est complexe ; le venin du mamba noir a révélé par une analyse en électrophorèse capillaire plus de 70 peptides dans la gamme de masse de 6000 à

9000 Da¹. Cette gamme de masse est caractéristique des neurotoxines de serpents dont certaines comme les dendrotoxines sont spécifiques des venins de mambas. Les venins de mambas contiennent également des fasciculines qui sont des inhibiteurs de l'acétylcholinestérase. Elles ont été nommées ainsi à cause de leur effet sur certains muscles, provoquant des contractions spontanées et irrégulières aussi appelées fasciculations. Leur effet est similaire à celui causé par certains gaz de combat, comme le gaz sarin, à cause de la production de fortes concentrations d'acétylcholine au niveau des synapses et de la jonction neuromusculaire.

D'autres types de toxines sont également représentés au sein des venins de mambas comme les toxines de type muscarinique (MT)^{2,3} que l'on retrouve aussi dans les venins de cobras⁴. La structure de ces toxines s'apparente à trois doigts d'une main humaine (*Figure 2*) et elles sont de ce fait appelées « toxines à trois doigts ».

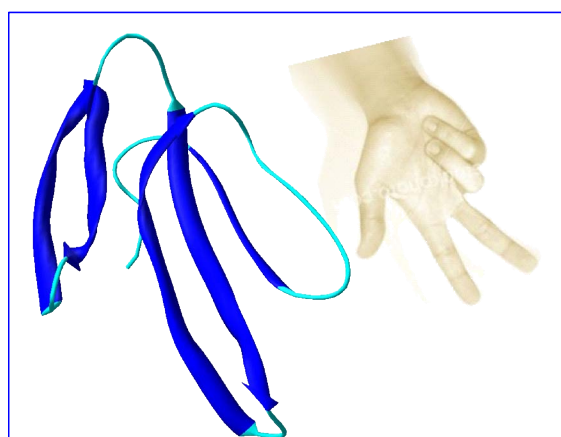


Figure 2: Toxine muscarinique MT2, une toxine de structure similaire à trois doigts

Parmi ces toxines à trois doigts, on trouve également les fasciculines et les α -neurotoxines. L'un des points surprenant concernant ces toxines à trois doigts est que de très faibles quantités de protéines endogènes de structures similaires aux toxines à trois doigts ont été récemment découvertes dans le système nerveux central des mammifères⁵. Cette observation nous ramène naturellement à la symétrie qui a été observée précédemment entre les sarafotoxines de serpents et les endothélines des mammifères.

Au contraire des toxines spécifiques des canaux ioniques trouvées généralement dans les venins de serpents, les toxines muscariniques ne ciblent pas les récepteurs ioniques mais plutôt les récepteurs couplés aux protéines G (RCPG) et plus précisément les récepteurs muscariniques dont le rôle principal est le contrôle de la tonicité de certains muscles lisses.

Les RCPG sont des protéines transmembranaires capables de reconnaître des messages aussi différents que la lumière, les odeurs, des hormones et des neurotransmetteurs. Il s'agit de la plus grande famille de récepteurs membranaires puisque plus d'un millier de RCPG différents sont aujourd'hui connus. Leur étude est de première importance : 50% des médicaments de la pharmacopée agissent directement sur ces récepteurs. Leur étude reste cependant très délicate puisqu'il s'agit de systèmes dynamiques très difficiles à purifier et quasiment impossibles à cristalliser. Il est donc primordial de trouver de nouveaux outils pharmacologiques spécifiques d'un type de RCPG donné afin d'en déterminer en détail la distribution cellulaire et les mécanismes d'actions. Seul un nombre très restreint de toxines actives sur les RCPG ont été découvertes. Ainsi, les conotoxines sont actuellement les composés les plus utilisés comme outils pharmacologiques pour l'étude des récepteurs adrénergiques (famille de RCPG).

La motivation principale de ces travaux est de découvrir, dans un venin de mamba, de nouvelles toxines spécifiques de récepteurs couplés à la protéine G dont aucun ligand n'est encore connu. Le venin du mamba que nous avons choisi, *Dendroaspis angusticeps*, a déjà été étudié au niveau moléculaire. Néanmoins, la complexité des venins de mambas comparée au peu de toxines qui ont été caractérisées permet d'envisager que de nombreux autres composés bioactifs n'aient pas encore été caractérisés. Dans cette optique, l'activité de ce venin a été évaluée sur différents récepteurs, dont les RCPG.

III. Stratégie utilisée pour caractériser de nouvelles activités dans le venin de *Dendroaspis angusticeps*

La stratégie employée pour étudier l'activité des différents constituants du venin de ce mamba vert est composée de nombreuses étapes (*Figure 3*).

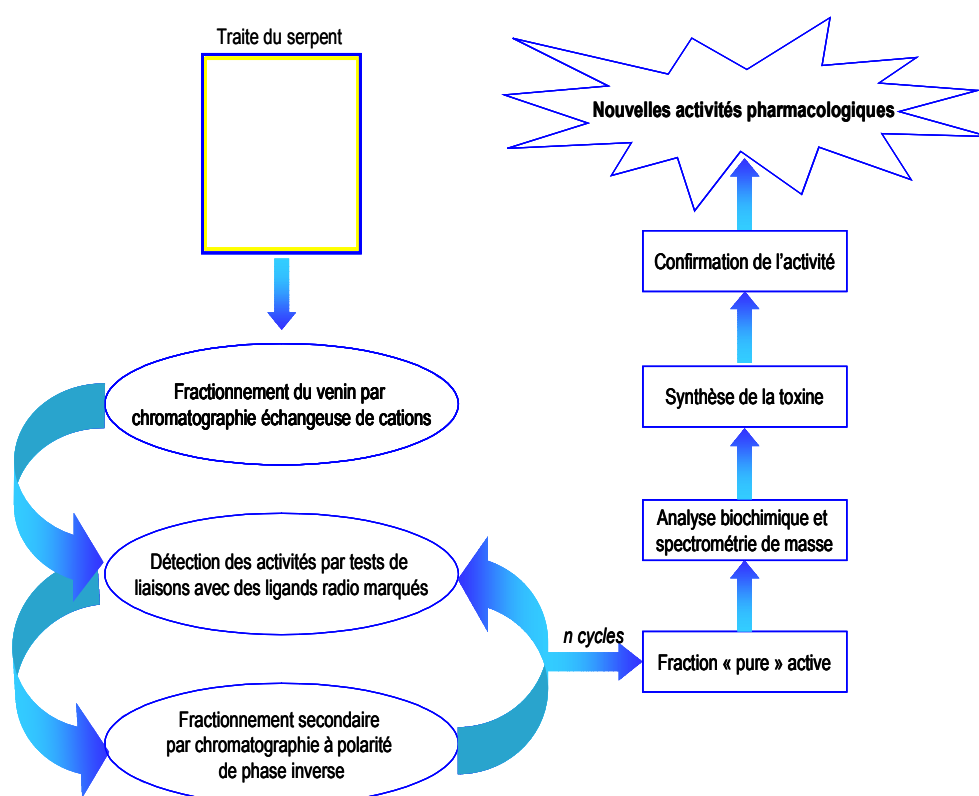


Figure 3: Stratégie de découverte de nouvelles activités pharmacologiques employée pour le venin de *Dendroaspis angusticeps*

La première étape est le criblage des différentes activités biologiques (par des tests de liaison avec des ligands radiomarqués) présentes dans le venin après son fractionnement. La toxine responsable de l'activité est alors isolée et sa caractérisation structurale est réalisée par une combinaison de méthodes biochimiques et de spectrométrie de masse. Une fois la toxine complètement caractérisée, sa synthèse est entreprise et des études d'activité plus approfondies sont réalisées.

A. Criblage des activités biologiques présentes dans du venin

L'étape initiale est la « traite » du serpent qui permet de prélever le venin brut à partir de l'animal vivant. Le venin, liquide jaunâtre plus ou moins visqueux selon les serpents, est recueilli dans un bécher (ou cristallisoir) et lyophilisé.

Etant donné le nombre important de constituants présents dans ces venins, une première étape de fractionnement par chromatographie échangeuse de cations est effectuée.

Les différents composés sont séparés par différence de charge globale puisque l'élution s'effectue par un gradient de chlorure de sodium. Les composés les plus acides (chargés plus

négativement) seront élués plus rapidement que les composés basiques chargés positivement. L'activité biologique de chacune des fractions obtenue est évaluée grâce à des tests de liaisons (*binding*) employant des ligands radioactifs marqués, comme cela a été décrit dans le premier chapitre.

Une fois qu'une activité est détectée, la fraction est sélectionnée puis de nouveau fractionnée, cette fois-ci par chromatographie de partage à polarité de phase inverse. Ce changement de mode de séparation, ou chromatographie bidimensionnelle « off-line », a pour but d'augmenter l'efficacité de la séparation. L'activité déterminée préalablement est recherchée dans ces nouvelles fractions. Plusieurs étapes de « purification chromatographique/tests de liaison » sont parfois nécessaires pour obtenir un pic unique, dans le chromatogramme, porteur de l'activité biologique.

B. Analyse structurale d'une toxine active purifiée

La structure primaire de la toxine d'intérêt est généralement déterminée en utilisant la dégradation d'Edman, des digestions enzymatiques et la spectrométrie de masse en MS et MS/MS. La dégradation d'Edman permet d'obtenir une partie des séquences recherchées mais nécessite à la fois un échantillon pur et en quantité importante. Au contraire, la spectrométrie de masse permet d'isoler l'espèce à fragmenter parmi les autres et est une technique extrêmement sensible qui requiert peu de matériel. L'approche « bottom-up » qui permet d'établir une cartographie peptidique de la toxine et de fragmenter sélectivement les peptides correspondant aux zones non séquencées par Edman (extrémité C-terminale), ajoute un intérêt considérable à cette technique pour l'analyse des toxines de haut poids moléculaire. Enfin l'approche innovante « top-down » consistant à fragmenter les protéines entières en phase gazeuse, est bien adaptée à l'étude des petites protéines et permet d'obtenir des informations de structure à partir de peu d'échantillon.

IV. L'AdTx1, une nouvelle toxine issue du venin *Dendroaspis angusticeps* ligand spécifique des récepteurs α_{1A} -adrénergiques

Le venin du *Dendroaspis angusticeps* est un venin très complexe. L'analyse chromatographique à polarité de phase inverse couplée avec le spectromètre FT-ICR permet de détecter plus de 150 composés différents (*Figure 4, Figure 5*).

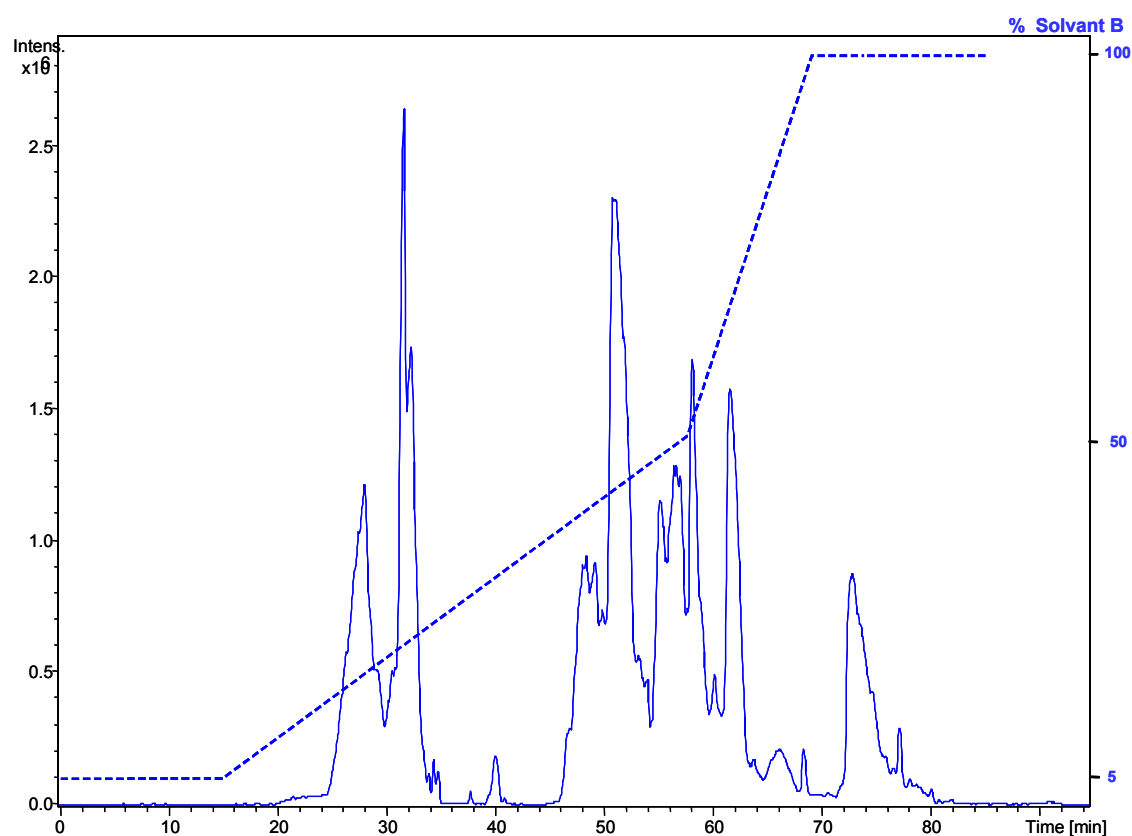


Figure 4: Chromatogramme obtenu par nanoLC-FT-ICR pour le venin de *Dendroaspis angusticeps*

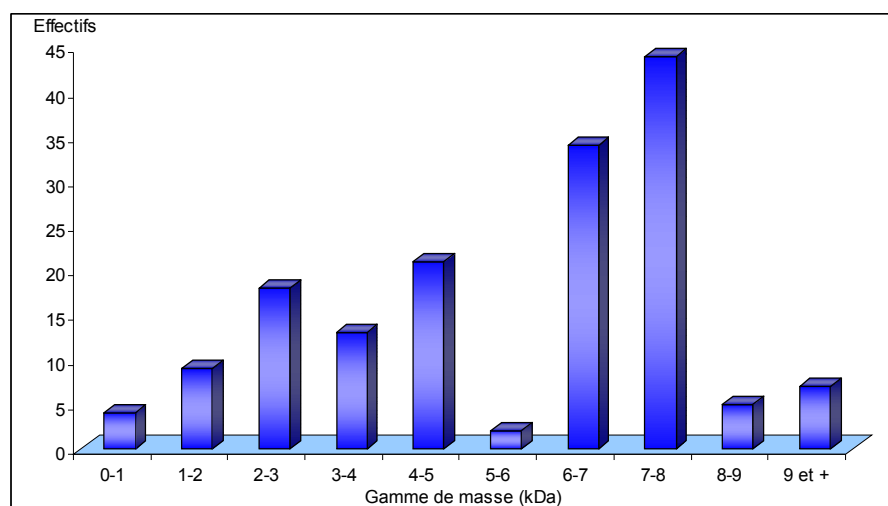


Figure 5: Répartition massique des toxines de *Dendroaspis angusticeps* détectées par nanoLC-FT-ICR

La répartition en masse des composants du venin, montre que la majorité des toxines ont des masses comprises entre 6000 et 8000 Da. Cette gamme de masse est tout à fait caractéristique des toxines qui ont été découvertes dans les venins de mambas comme les

toxines muscariniques, les fasciculines ou les dendrotoxines. Une douzaine d'entre elles ont déjà été étudiées mais la complexité de ce venin, démontrée par cette simple analyse, confirme le fait que de nombreuses autres toxines actives n'ont pas encore été caractérisées.

A. Purification d'une toxine active sur les récepteurs α_1 -adrénergiques

Pour la recherche d'activité, le venin brut de *Dendroaspis angusticeps* est fractionné par chromatographie échangeuse de cations. Les composés sont élués par un gradient de chlorure de sodium à 1 mol.L⁻¹. Le chromatogramme obtenu, assez complexe, permet de recueillir 13 fractions identifiées A à M (Figure 6).

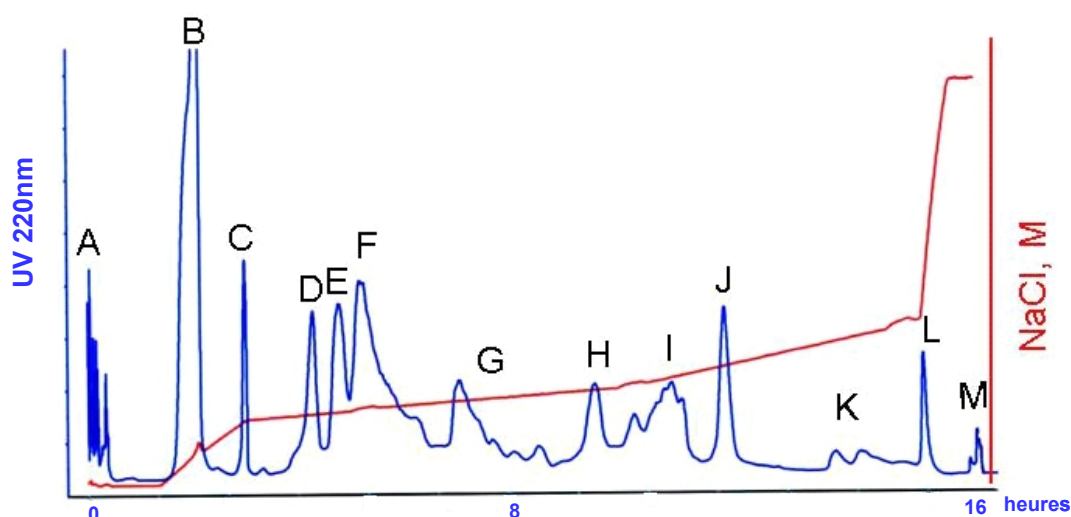


Figure 6: Fractionnement du venin brut par chromatographie échangeuse de cations – Les fractions collectées sont indiquées directement sur le chromatogramme

La nomenclature de chacune de ces fractions suit le modèle « DA₁X », dans lequel DA₁ signifie que l'échantillon d'étude est le premier de *Dendroaspis angusticeps* et où la lettre X caractérise la lettre attribuée à la fraction chromatographique, donc variant de A à M. Chacune de ces fractions est alors testée sur différents modèles récepteurs/ligands. Pour les récepteurs α_1 -adrénergiques, le modèle employé est le synaptosome de rat mis au contact de Prazosine-³H (ligand tritié). Le synaptosome de rat contient un très grand nombre de récepteurs différents dont des récepteurs adrénérgiques qui disposent de nombreux sous-types

(Figure 7). Aucune sélection particulière des récepteurs α_1 -adrénergiques n'est obtenue à ce niveau.

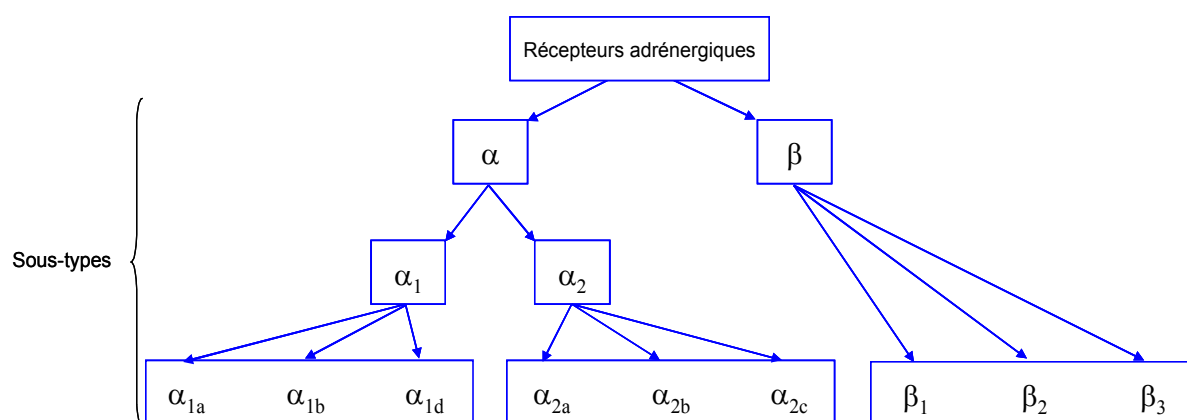


Figure 7 : Les récepteurs adrénérgiques sont composés de nombreux sous types

En revanche, l'avantage de la prazosine est qu'elle présente une très bonne sélectivité pour les récepteurs adrénérgiques de type α_1 . Bien qu'elle permette de s'intéresser uniquement aux α_1 , elle ne permet pas à ce niveau de distinguer la sélectivité du ligand pour les différents sous-types α_{1A} , α_{1B} et α_{1C} .

D'un point de vue expérimental, le synaptosome, la Prazosine- ^3H et un volume de fraction issue du venin sont ajoutés au même instant. Un délai d'une heure environ permettant d'atteindre l'équilibre du système est imposé. L'excès de ligand tritié non fixé sur les récepteurs est éliminé par filtration et lavage. La radioactivité du synaptosome est alors mesurée. Comme expliqué dans le premier chapitre, cette mesure de radioactivité brute est comparée à deux références : le NS (*Non-specific*) et le BT (*Binding Total*). Les résultats obtenus sur les fractions DA1A à DA1M (Figure 8) montrent deux cas de figure.

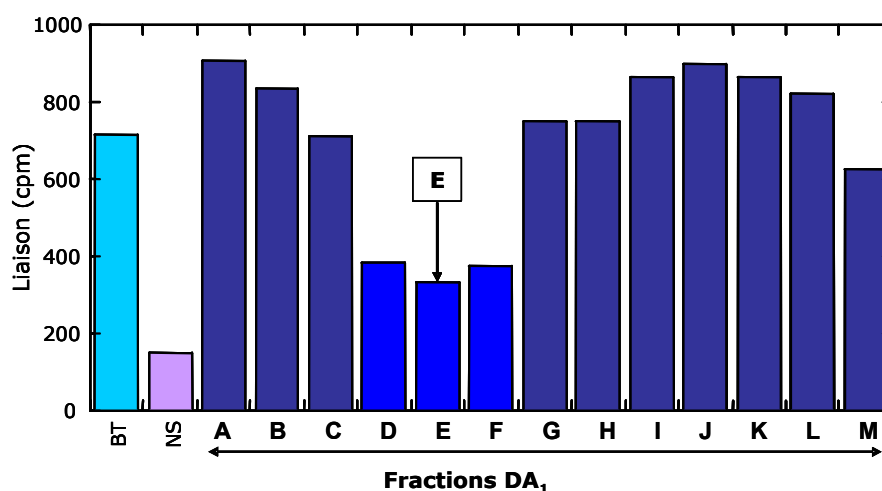


Figure 8: Expériences de liaison des fractions de DA1A à DA1M sur les récepteurs α_1 -adrénergiques en présence de 3H-Prazosine - BT= Binding Total, NS= Liaison non spécifique

Tout d'abord, pour les fractions DA1A à C et DA1G à M, la radioactivité mesurée est proche de la valeur de la liaison totale. Ceci signifie qu'un maximum de Prazosine- ^3H est resté fixé aux récepteur α_1 et que, de ce fait, aucun composé n'est venu perturber cette liaison. Une autre explication pourrait être qu'un composé se fixe bien aux récepteurs α_1 mais que la variation de signal obtenue n'est pas significative (composés de faible concentration). Un autre cas de figure est visible pour les fractions DA1D, E et F, puisque le signal mesuré décroît fortement. Cette baisse d'intensité traduit le fait qu'une toxine a fortement perturbé la liaison Prazosine- ^3H /récepteur. Ces fractions contiennent de nombreux composés et un nouveau fractionnement est nécessaire. Dans la suite de ce travail, seule la fraction DA1E sera considérée.

Le chromatogramme de séparation par chromatographie à polarité de phase inverse (RP-HPLC, détection UV) de cette fraction est présenté sur la *Figure 9*. La trace UV complexe, traduit la séparation de nombreux composés et, de ce fait, l'efficacité de la chromatographie bidimensionnelle. Vingt fractions sont recueillies et nommées DA1EA à DA1ET, la lettre ajoutée informant sur le nom de la fraction issue de la seconde séparation.

Afin de retrouver l'activité sur les récepteurs α_1 -adrénergiques, chacune de ces fractions secondaires a été testée comme précédemment en utilisant les synaptosomes de rat et la prazosine tritiée. Les résultats obtenus pour les 20 fractions sont regroupés *Figure 10*.

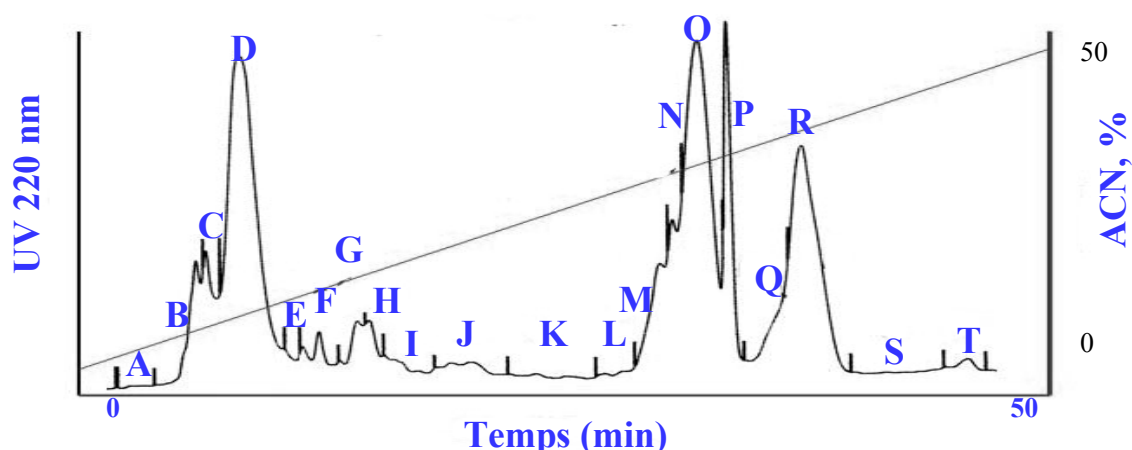


Figure 9: Chromatogramme de fractionnement de DA₁E par RP-HPLC

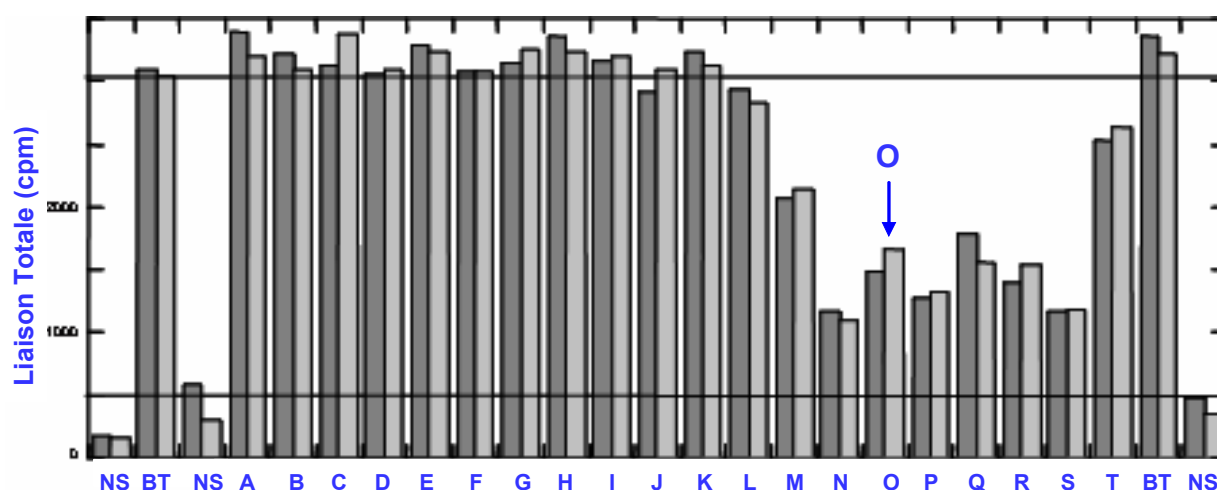


Figure 10: Expériences de liaison des fractions de DA₁EA à DA₁ET sur les récepteurs α_1 -adrénergiques en présence de 3H-Prazosine – BT= *Binding Total*, NS= *Liaison non spécifique*, BF= *radioactivité résiduelle piégée par le filtre*

Toutes ces expériences ont été doublées pour confirmer les variations de signal observées. Les fractions DA₁EA à DA₁EL ne présentent pas de variations significatives et ne contiennent donc pas la (ou les) toxine(s) actives sur les récepteurs étudiés. En revanche, une baisse significative de la radioactivité est observée pour les sept fractions allant de DA₁EM à DA₁ET. La liaison de la prazosine tritiée et des récepteurs α_1 -adrénergiques est perturbée par une ou des toxines qui se lient aux mêmes récepteurs. Une seule fraction est sélectionnée pour continuer notre étude, la fraction DA₁EO.

Cette fraction est une dernière fois purifiée (système RP-HPLC analytique) et les nouvelles fractions testées biologiquement sur le même modèle afin d'obtenir la fraction DA₁EOA

active et présentant un unique pic chromatographique (dans les conditions de séparation employées). Une analyse par spectrométrie de masse est ensuite entreprise.

B. Détermination de la séquence primaire de la toxine

B.1. Mesure de masse précise et réduction des ponts disulfures

Une solution de concentration théorique à $4 \text{ pmol} \cdot \mu\text{L}^{-1}$ est analysée par nanoESI-FT-ICR (*Figure 11*).

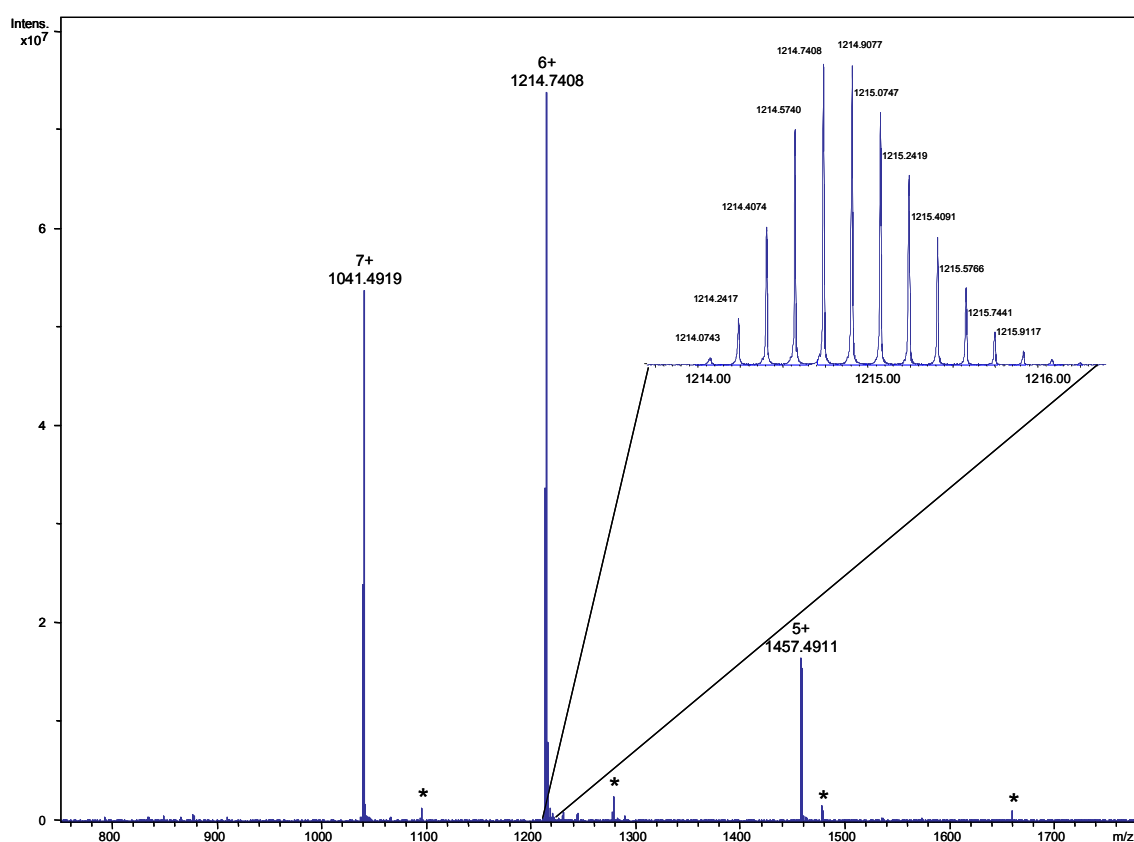


Figure 11: Spectre nanoESI-FT-ICR de la toxine native DA1EOA du venin du mamba vert – * indique les impuretés présentes

Malgré la présence de quelques impuretés minoritaires, un composé très majoritaire est observé. La masse monoisotopique pour cette toxine est mesurée à $M_{\text{ox}} = 7278,3851 \text{ Da}$ (Ox signifie que les ponts S-S sont oxydés)

Dans un second temps, les ponts disulfures de la toxine ont été réduits. La solution obtenue, purifiée sur Zip-Tip® C₁₈ est directement analysée par nanoESI-FT-ICR. Le spectre obtenu (Figure 12) présente une large distribution d'états de charge allant de l'espèce $[M+5H]^{5+}$ à l'espèce $[M+11H]^{11+}$. Cette allure est tout à fait celle d'une protéine « dépliée », pour laquelle de nouveaux sites de protonation sont accessibles.

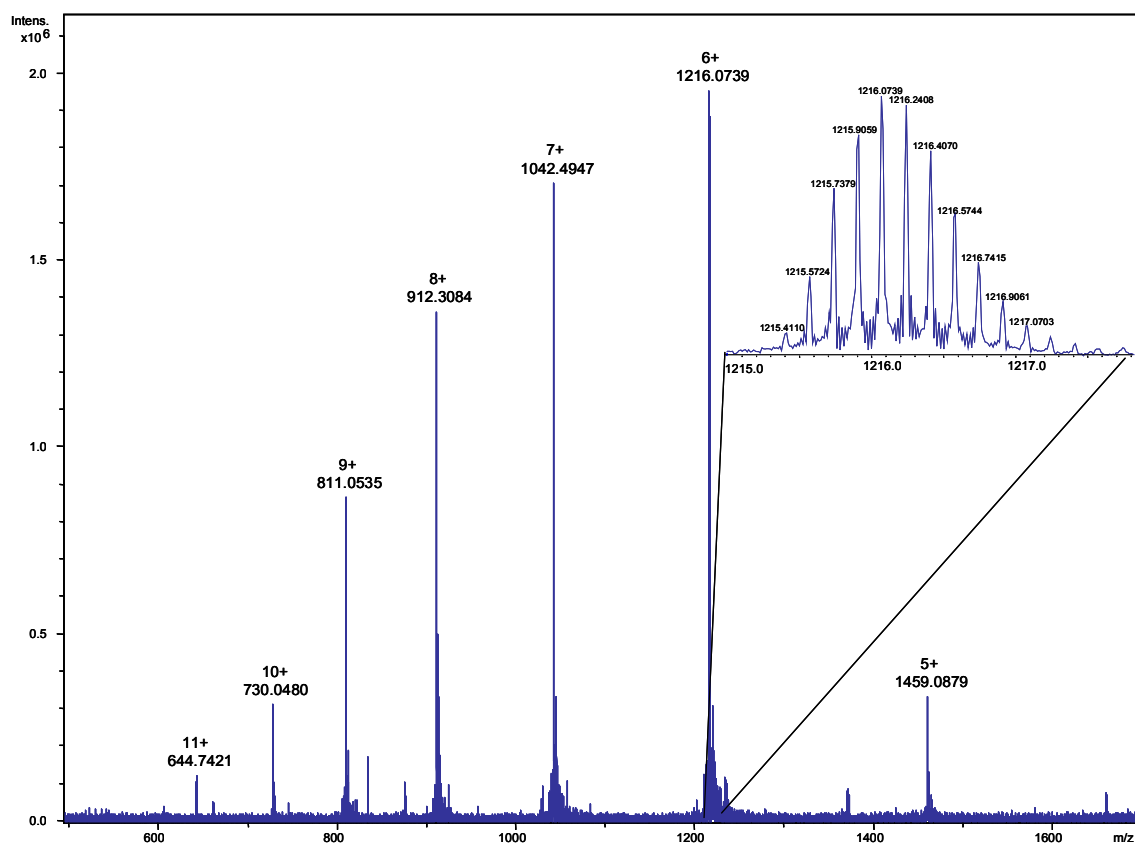


Figure 12 : Spectre nanoESI-FT-ICR de la toxine DA1EOA après réduction des ponts S-S – 10 scans

La masse mesurée après l'étape de réduction est $M_{\text{red}} = 7286,3541$ Da. La Figure 13 présente un agrandissement des massifs isotopiques des espèces $[M+6H]^{6+}$ de la toxine (A) avant et (B) après l'étape de réduction. L'écart mesuré entre les deux massifs est de 1,3367 Da. En tenant compte du fait que le massif correspond à un ion chargé 6 fois, on obtient une différence de masse avant/après réduction de 8,02 Da.

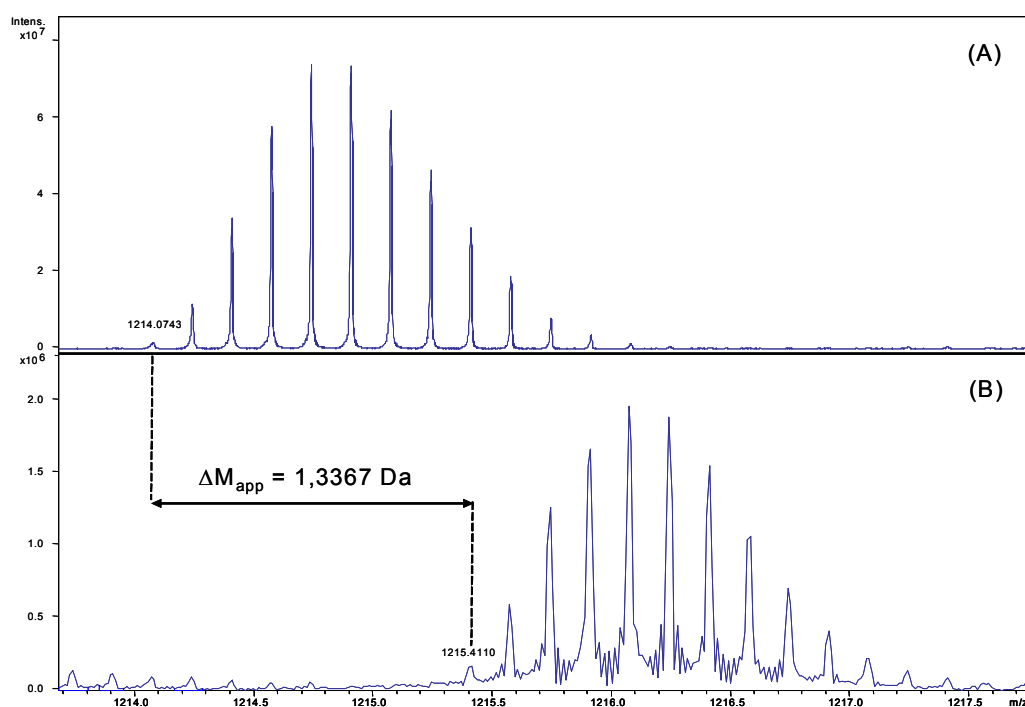


Figure 13: Zoom sur le massif isotopique de l'ion $[M+6H]^{6+}$ de la toxine (A) avant réduction et (B) après réduction

Cet écart caractérise la présence de 4 ponts disulfures dans la séquence primaire de la toxine. La gamme de masse ainsi que le nombre de ponts observés sont caractéristiques des toxines muscariniques et des fasciculines. Un séquençage d'Edman de la toxine a ensuite été entrepris.

B.2. Séquençage d'Edman

La masse de la toxine mesurée est de $M_{ox} = 7278,3851$ Da, ce qui correspond à une toxine d'environ 66 acides aminés (en comptant 110 Da de moyenne par acide aminé). La dégradation d'Edman de la toxine a permis d'obtenir les 50 premiers acides aminés (Figure 14) mais n'a pas pu fournir l'ensemble de la séquence primaire de la toxine en raison de différentes limitations expérimentales (voir annexe).

LTXVT SKSIF GITTE DXPDG QNLXF KRRHY VVPKI YDSTR GXAAT XPIPE xxxxxx

Figure 14: Les 50 premiers résidus de la toxine obtenus par dégradation d'Edman – les « X » indiquent un trou dans l'analyse caractérisant souvent une cystéine engagée dans un pont disulfure – les « x » indiquent que la séquence est incomplète en extrémité C-terminale.

La masse de cette séquence incomplète est de 5532,6761 (réduite) permettant de déterminer qu'il « manque » une masse 1753,6780 Da pour la compléter (une quinzaine d'acides aminés). Malgré une mesure de masse très précise, il n'est pas possible de déterminer la composition en acides aminés d'une masse aussi élevée. Néanmoins, la séquence incomplète permet d'effectuer une recherche par homologie de séquence (BLAST) dans les bases de données. Cette étape permettra de rechercher la présence de composés qui possèdent une séquence similaire à DA₁EOA (Tableau 1).

Nom de la toxine	Serpent	Séquence	M _{red} (Da)
Toxine recherchée	DA	LTCTVTSKSIF GITTEDCPDG QNLCTFKRRHY VVPKIYDSTR GCAATCPIPE XXXXXXXX XXXXX	7286,3541
Muscarinic toxin Beta	DPP	LTCVTSKSIF GITTEDCPDG QNLCTFKRRHY VVPKIYDITR GCVATCPIPE NYDSIHCCCT DKCNE	7340,4452
Synergistic-like venom protein	DPP	LTCVTSKSIF GITTEDCPDG QNLCTFKRRHY VVPKIYDITR GCVATCPIPE NYDSIHCCCT EKCNN	7339,4612
Muscarinic toxin-like protein 2	NK	LTCVKEKSIF GVTTEDCPDG QNLCTFKRWHM IVPGRYKKTR GCAATCPIAE NRDVIECCST DKCNL	7293,4704
Muscarinic toxin 3	DA	LTCVTKNITF GITTENCPAG QNLCTFKRWHY VIPRYTEITR GCAATCPIPE NYDSIHCCCT DKCNE	7382,4459
Muscarinic toxin 4	DA	LTCVTSKSIF GITTENCPDG QNLCTFKKWYY IVPRYSDITW GCAATCPKPT NVRETIHCCET DKCNE	7493,4555
Muscarinic toxin 1	DA	LTCVTSKSIF GITTENCPDG QNLCTFKKWYY IVPRYSDITW GCAATCPKPT NVRETIHCCET DKCNE	7512,4977

Tableau 1 : Alignement des 50 premiers résidus de DA₁EOA avec les séquences des 6 toxines les plus proches obtenues par BLAST- Les cystéines et les acides aminés variant sont colorés respectivement en rouge et en bleu - DA= *Dendroaspis angusticeps*, DPP= *Dendroaspis polylepsis polylepsis* et NK= *Naja kaouthia*)

Les résultats montrent que les séquences les plus proches de DA₁EOA dans les bases de données correspondent essentiellement à des toxines muscariniques. Toutes les séquences trouvées sont issues de venins de mambas sauf l'une d'elles qui provient de *Naja kaouthia*, un cobra. Les cobras et les mambas font partis de la même famille, les *Elapidae*. La similarité structurale des toxines observées dans les venins de ces deux familles s'explique par la présence d'un ancêtre commun à ces deux serpents⁶.

Pour déterminer la séquence complète de cette toxine, deux approches basées sur des expériences de spectrométrie de masse FT-ICR ont été utilisées et comparées.

B.3. Séquençage de l'extrémité C-terminale de DA₁EOA par une approche « bottom-up »

Cette approche permet non seulement d'effectuer une carte peptidique précise et, de ce fait, de confirmer la séquence obtenue par la dégradation d'Edman, mais elle permet aussi de mener des expériences de fragmentations spécifiques de peptides caractéristiques de l'extrémité C-terminale de la toxine.

Pour cette expérience, environ 1,4 nmol (soit 10 μ g) de la toxine DA₁EOA sont digérés par 27 pmol de trypsine dans un tampon de bicarbonate d'ammonium (pH=7,9). Le rapport trypsine/toxine, classique, est de l'ordre de 1/50. La trypsine, qui permet le clivage en position C-terminale des deux acides aminés arginine et lysine, est choisie car la dégradation d'Edman a révélé la présence de nombreux résidus de ce type. Après digestion, la toxine est réduite par un excès de TCEP. Le fait de n'avoir ni réduit, ni alkylé la toxine avant la digestion permet d'obtenir des conditions de digestion ménagées puisque certains sites de clivages ne sont pas accessibles à l'enzyme. Ces conditions ont pour but d'obtenir une carte peptidique simple, composée de peu de peptides. Environ 50 pmol du mélange de peptides réduits sont purifiés sur ZipTip[®] C₁₈. Les peptides sont élués directement par 10 μ L de solvant d'électrospray (H₂O/ACN/AF, 49,95/ 49,95/ 0,1%) et analysés par nanoESI-FT-ICR (*Figure 15*).

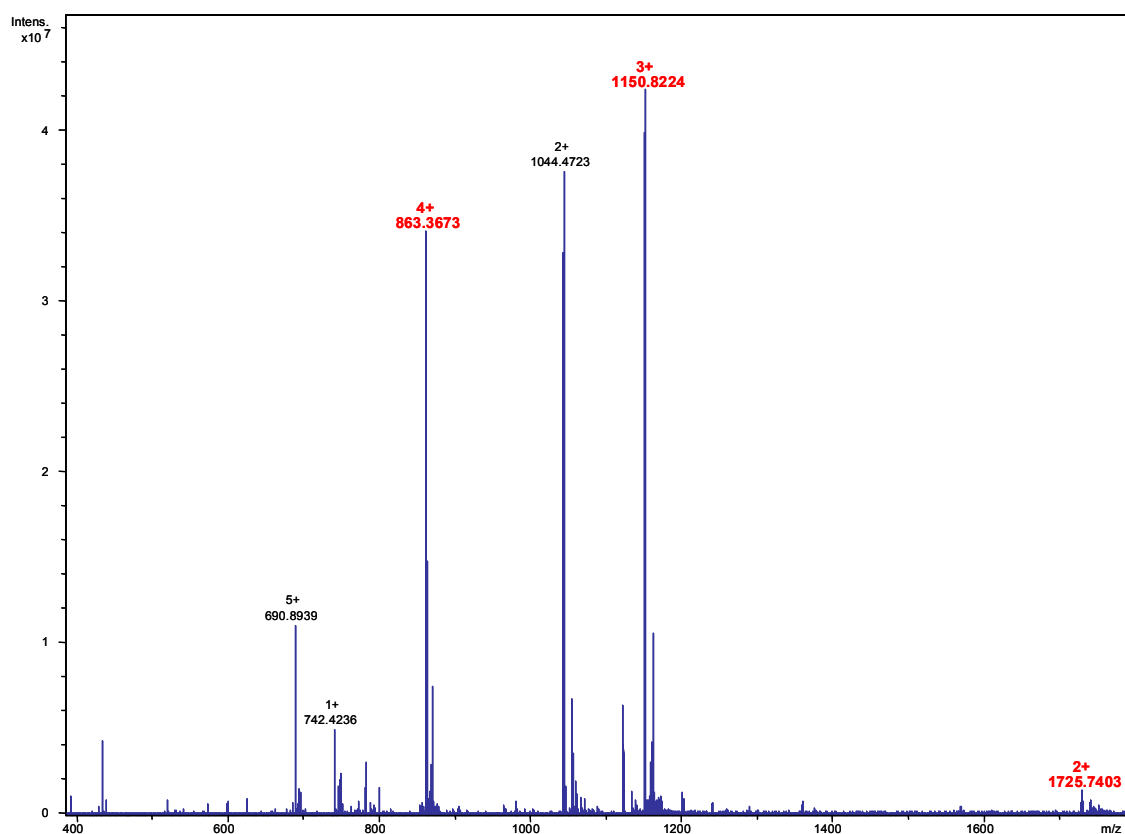


Figure 15: Spectre nanoESI –FT-ICR du digestat trypsique de DA₁EOA

Comme attendu, le spectre de masse obtenu est assez simple. Certains peptides sont identifiés rapidement grâce à la séquence incomplète et à la précision en masse de l'instrument. Ces peptides permettent d'étalonner notre spectre en interne (*Figure 16*) de

manière à obtenir une mesure de masse très précise sur les ions non attribués (précision de l'ordre de 1 ppm).

Les signaux les plus intenses (en rouge sur le spectre) sont les espèces $[M+2H]^{2+}$, $[M+3H]^{3+}$ et $[M+4H]^{4+}$ d'un peptide de 3449,4454 Da. Cette masse ne pouvant être reliée à une séquence déterminée précédemment par dégradation d'Edman, on peut supposer que la séquence correspondante contient une partie (voire la totalité) de l'extrémité C-terminale manquante.

Peptides retrouvés:

M_{exp} (Da)	Séquence
741,4163	HYVVPK
750,3937	LTCVTSK
2086,9300	SIFGITTEDCPDGQNLCKFK
2243,0340	SIFGITTEDCPDGQNLCKFKR
2399,1313	SIFGITTEDCPDGQNLCKFKRR
3449,4454	Peptide C-terminal?

Couverture de séquence:

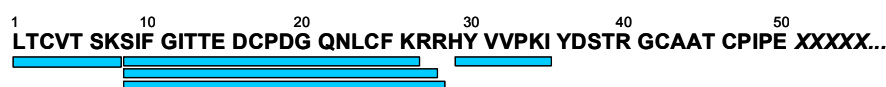
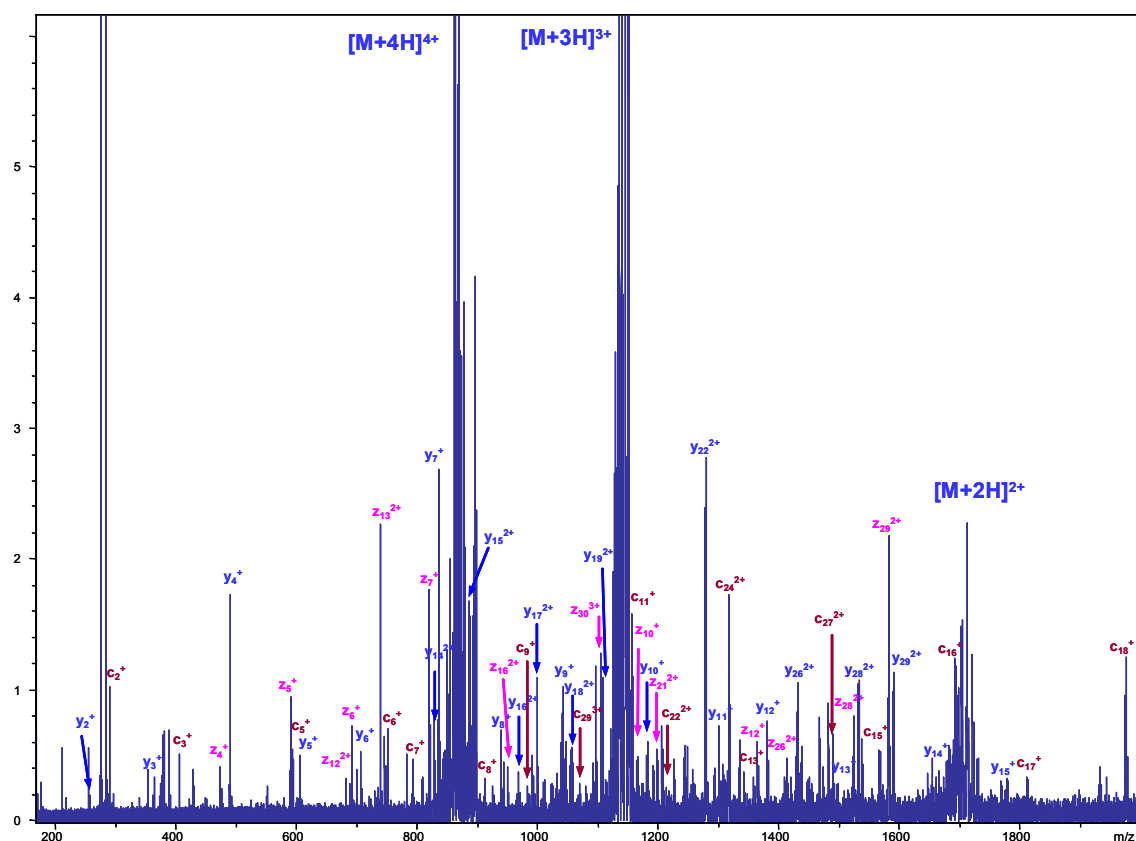


Figure 16: Cartographie peptidique de DA₁EOA

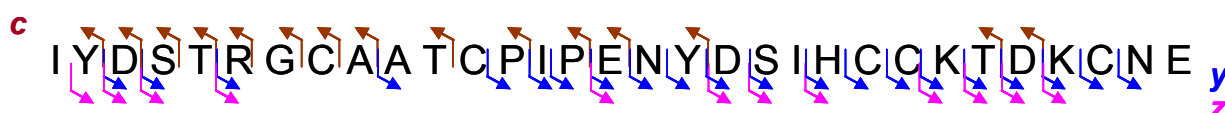
Des expériences de MS/MS ont donc été menées sur l'ion (4^+) à m/z 863,8671. La fragmentation par collision avec du xénon (SORI-CAD) conduit à peu de fragments ne permettant pas d'obtenir des informations importantes sur la séquence du peptide. En revanche, la fragmentation du même peptide par ECD donne un spectre beaucoup plus complexe et de ce fait, beaucoup plus informatif (Figure 17).

L'analyse de ce spectre indique la présence de nombreux fragments c et z , comme attendu pour des expériences d'ECD mais aussi de fragments y . La présence de ces ions y , caractéristiques des fragmentations de basse énergie avec un gaz inerte, s'explique sans doute par une légère excitation vibrationnelle des ions (sans doute au moment de la sélection des précurseurs). Néanmoins, le séquençage *de novo* du peptide se révèle délicat car les ions c et z sont parfois décalés de 1 voir 2 Da par rapport à leur masse théorique. Ces décalages, couramment observés lors des expériences ECD sont la conséquence de transferts d'hydrogènes en phase gazeuse après la capture de l'électron (et probablement la rupture de la liaison N-C α).



**Figure 17: Spectre ECD de l'ion (4^+) à m/z 863,8671 – temps d'irradiation par des électrons (~ 1 eV/250 μ s)
– 70 accumulations**

Malgré cette complexité, la résolution et la précision du spectromètre de masse FT-ICR permettent de déterminer la séquence de ce peptide mais trois ambiguïtés mineures sont néanmoins visibles sur la séquence. Il s'agit de trois doublets pour lesquels aucune fragmentation n'est obtenue : « AT », « SI » et « NE ». La précision en masse permet de déterminer les bons acides aminés mais leur enchaînement reste incertain. Ces ambiguïtés seront levées lors de la comparaison de cette séquence avec les autres toxines connues par BLAST.



Grâce à la calibration interne, la masse expérimentale du peptide d'intérêt est mesurée à 3449,4454 Da tandis que la masse théorique du peptide proposée est de 3449,4455 Da. Cette excellente précision ($\Delta M < 0,3$ ppm) montre toute l'utilité d'une calibration interne en FT-ICR.

A titre de comparaison, la *Figure 18* présente le spectre obtenu pour la fragmentation de la toxine par SORI-CAD. Le peu de fragments obtenus dans ce cas met en évidence l'efficacité de l'ECD.

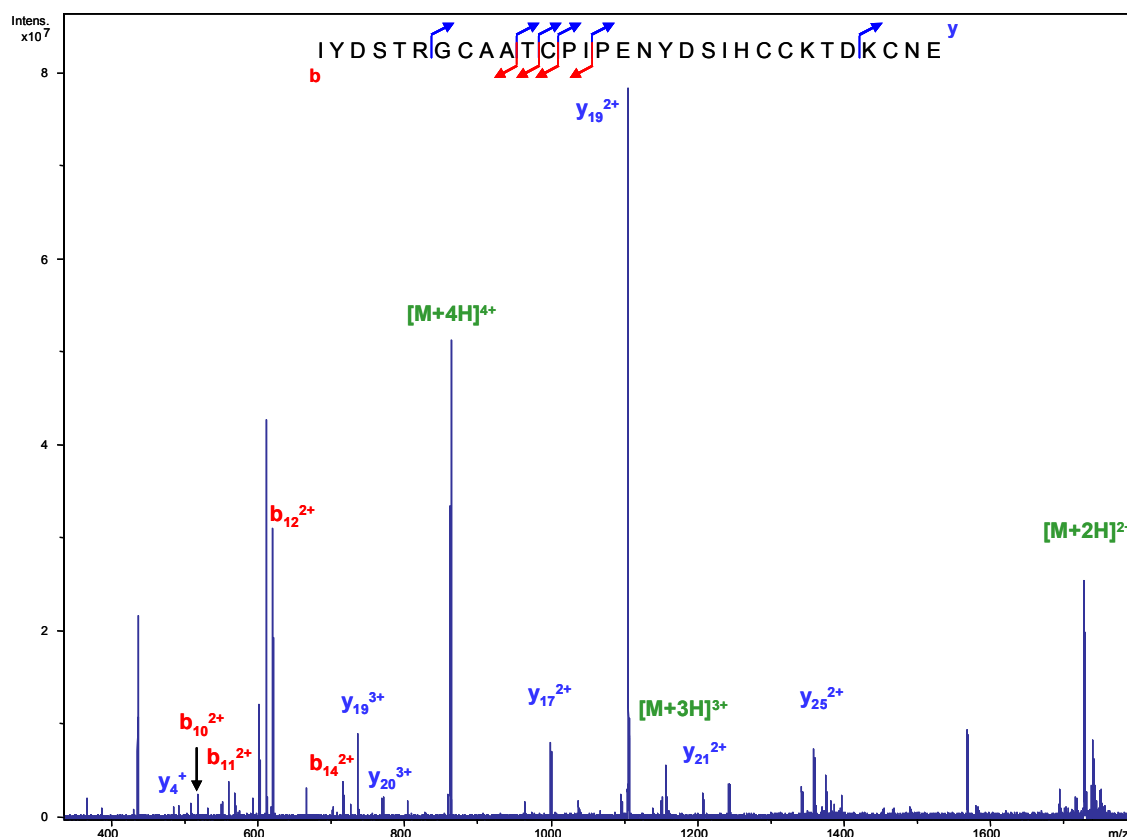


Figure 18: Spectre SORI-CAD de l'ion (4⁺) du peptide C-terminal de DA₁EOA

Il apparaît de nouveau que les fragmentations du côté *N*-terminal des prolines sont favorisées conduisant à la formation des ions fragments (dichargés) les plus intenses : b_{12}/y_{19} et b_{14}/y_{17} . Ces résultats montrent que l'ECD est, dans ce cas, bien plus efficace que le SORI-CAD. Ce phénomène ne peut être généralisé puisque d'autres toxines comme les sarafotoxines d'*Atractaspis irregularis* se fragmentent mieux par SORI-CAD que par ECD.

La séquence de la toxine entière est de nouveau comparée aux toxines trouvées par BLAST. Il apparaît que séquence de la toxine muscarinique *beta* du mamba noir *Dendroaspis polylepsis polylepsis* est très voisine de celle de DA₁EOA puisque seuls deux acides aminés les différencient: substitution d'une sérine par une isoleucine en position +38, et d'une alanine par une valine en position +43. Leurs extrémités *C*-terminales sont totalement identiques (*Tableau 2*).

Nom de la toxine	Serpent	Séquence								M_{red} (Da)
Toxine DA ₁ EOA	DA	LTCVTSKSIF	GTTTDCPDG	QNLCKRRHY	VVPKIYDSTR	GCAATCPIPE	NYDSIHCKT	DKCNE		7286,3541
Muscarinic toxin Beta	DPP	LTCVTSKSIF	GTTTDCPDG	QNLCKRRHY	VVPKIYDTR	GCVATCPIPE	NYDSIHCKT	DKCNE		7340,4452
Synergistic-like venom protein	DPP	LTCVTSKSIF	GTTTDCPDG	QNLCKRRHY	VVPKIYDTR	GCVATCPIPE	NYDSIHCKT	EKCNN		7339,4612
Muscarinic toxin-like protein 2	NK	LTCVKEKSIF	GVTTEDCPDG	QNLCKRWHM	IVPGRYKKTR	GCAATCPIAE	NRDVIECCST	DKCNL		7293,4704
Muscarinic toxin 3	DA	LTCVTKNITF	GTTTENCAG	QNLCKRWHY	VIPRYTETR	GCAATCPIPE	NYDSIHCKT	DKCNE		7382,4459
Muscarinic toxin 4	DA	LTCVTSKSIF	GTTTENCAG	QNLCKKWYY	IVPRYSBITW	GCAATCPKPT	NVRETIHCCET	DKCNE		7493,4555
Muscarinic toxin 1	DA	LTCVTSKSIF	GTTTENCAG	QNLCKKWYY	IVPRYSBITW	GCAATCPKPT	NVRETIHCCET	DKCNE		7512,4977

Tableau 2: Alignement de séquences entre Da1ENA et celles de différentes toxines muscariniques

Cette comparaison permet de confirmer que l'enchaînement des acides aminés au sein des trois doublets « AT », « SI » et « NE » puisqu'ils sont très conservés au sein des familles de toxines muscariniques.

La masse mesurée expérimentalement (7286,3741 Da) a été comparée à la masse théorique (7286,3619) : un écart de 1,67 ppm peut être calculé, ce qui est excellent dans cette gamme de masse.

B.4. Approche « top-down » par ECD sur la toxine entière

Bien que la séquence complète ait pu être déterminée par une approche classique, des expériences de fragmentation sur la toxine entière (réduite) en phase gazeuse ont été menées. Ces études ont pour but d'évaluer l'intérêt et les limitations de l'approche « top-down » sur protéine entière pour ce type d'application.

Compte tenu de la pureté de la toxine DA₁EOA dans sa fraction, aucune sélection d'ions parents n'a été réalisée avant irradiation par des électrons. Ainsi, l'ensemble des ions multichargés a été fragmenté par ECD. Supprimer l'étape de sélection permet d'avoir un signal très intense au début de l'analyse, condition nécessaire pour obtenir un spectre ECD de bonne qualité. Le spectre obtenu est présenté sur la *Figure 19*.

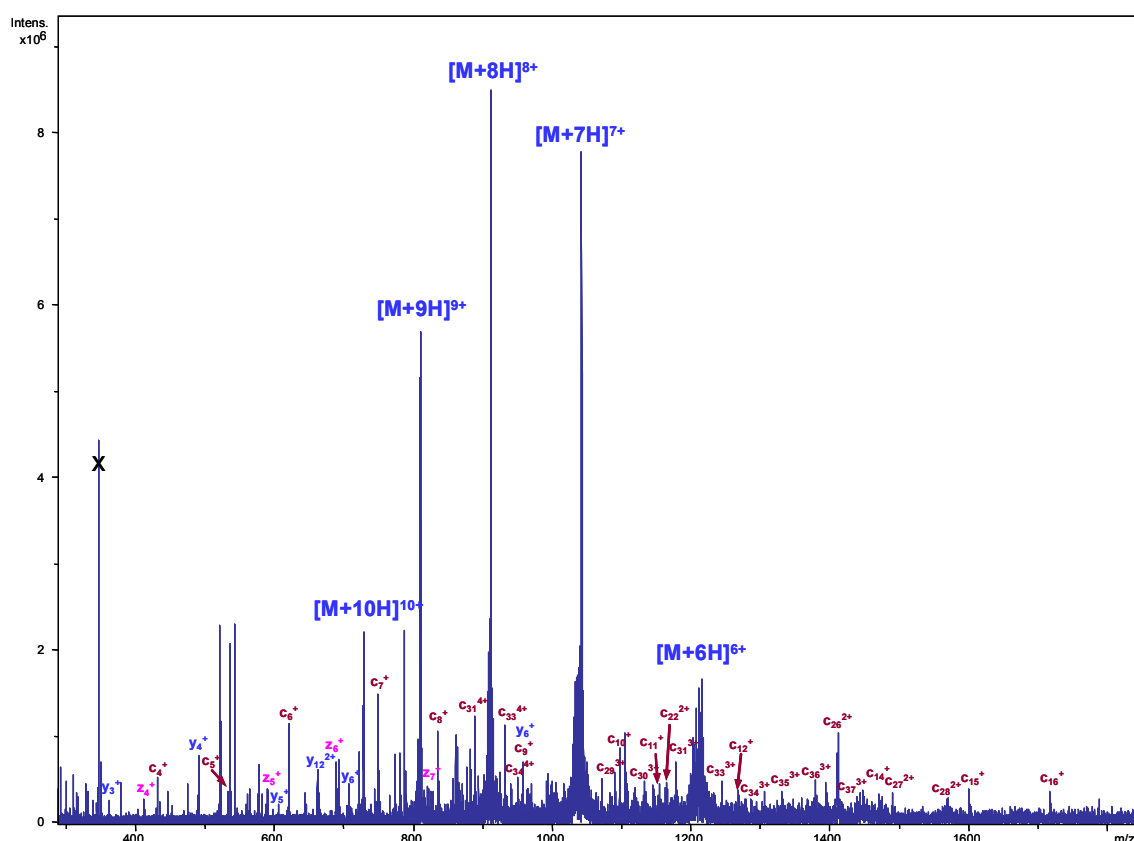
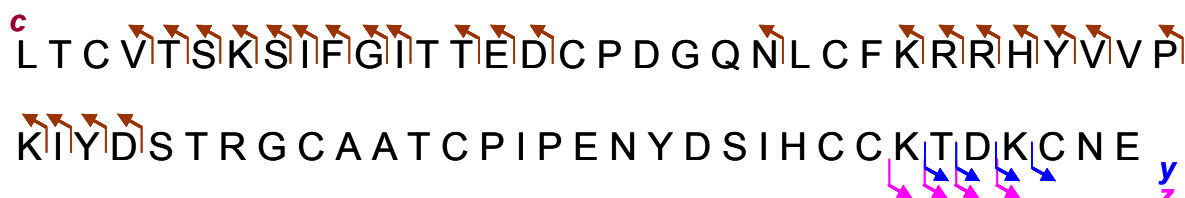


Figure 19: Spectre ECD de DA₁EOA - Irradiation 250ms par des électrons de ~1eV - 150 scans



Le spectre, de nouveau très complexe, permet l'attribution de nombreux fragments. En effet, une très belle série d'ions c permet de déterminer de longs tags de séquence. Quelques ions z et y permettent de déterminer les sept acides aminés C -terminaux. Il apparaît qu'aucun ion caractéristique n'est retrouvé pour la partie centrale de la toxine, ce qui indique probablement une structuration de la petite protéine en phase gazeuse, malgré la réduction des ponts disulfures. Cette structuration en phase gazeuse correspond à la formation de liaisons de faible énergie (type liaisons hydrogène), qui ne sont pas rompues par l'ECD. Pour pallier ce problème et obtenir des fragmentations sur toute la chaîne peptidique, il faudrait rompre ces interactions de faible énergie en phase gazeuse. Cette activation douce peut être réalisée par l'excitation des ions dans la cellule par des photons infrarouges de faible énergie. L'ajout d'un laser infrarouge est en projet sur le spectromètre de masse FT-ICR du laboratoire.

Ainsi, bien que l'approche « top-down » n'ait pas permis le séquençage complet de DA₁EOA, sa rapidité de mise en œuvre, comparée à l'approche « bottom-up » impliquant une digestion enzymatique, est très intéressante pour l'obtention rapide de données structurales partielles. A titre d'exemple, les tags obtenus grâce à l'ECD permettent dans notre cas d'obtenir les mêmes résultats de recherche par homologie de séquence qu'avec les données issues de dégradation d'Edman. Une autre application des ces données structurales est que les tags obtenus, assez longs, peuvent permettre la fabrication de sondes nucléotidiques spécifiques de cette toxine voire de cette famille de toxines. Ces sondes nucléotidiques, comme nous l'avons vu dans le cas des sarafotoxines, pourraient être utilisées pour sélectionner et amplifier uniquement les ARN messagers qui codent pour ces toxines.

La donnée manquante est le type d'organisation des ponts disulfures de DA₁EOA. Etant donné les grandes similitudes observées entre la séquence trouvée et celles des toxines muscariniques à trois doigts (notamment la MT β qui ne présente que deux acides aminés variants), nous considérerons par similarité que cette toxine adopte le même arrangement, à savoir : Cys₃-Cys₂₄, Cys₁₇-Cys₄₂, Cys₄₆-Cys₅₇ et Cys₅₈-Cys₆₃.

C. Recherche de la sélectivité de l'AdTx1 pour les sous-types du récepteur α_1 -adrénergique

La séquence de la toxine DA₁EOA ayant été déterminée, cette dernière a été synthétisée en quantités importantes afin de pouvoir réaliser plusieurs tests permettant de déterminer plus précisément son activité biologique. Pour ne pas confondre la toxine native avec la toxine synthétique, cette dernière est baptisée AdTX1 pour toxine adrénérergique 1.

Afin de confirmer l'activité de l'AdTx1, de nouvelles expériences de liaisons ont été menées. Elles consistent à mesurer le déplacement de la liaison à l'équilibre d'une concentration fixe de ³H-prazosine en présence d'une concentration variable et croissante d'AdTx1 (*Figure 20*).

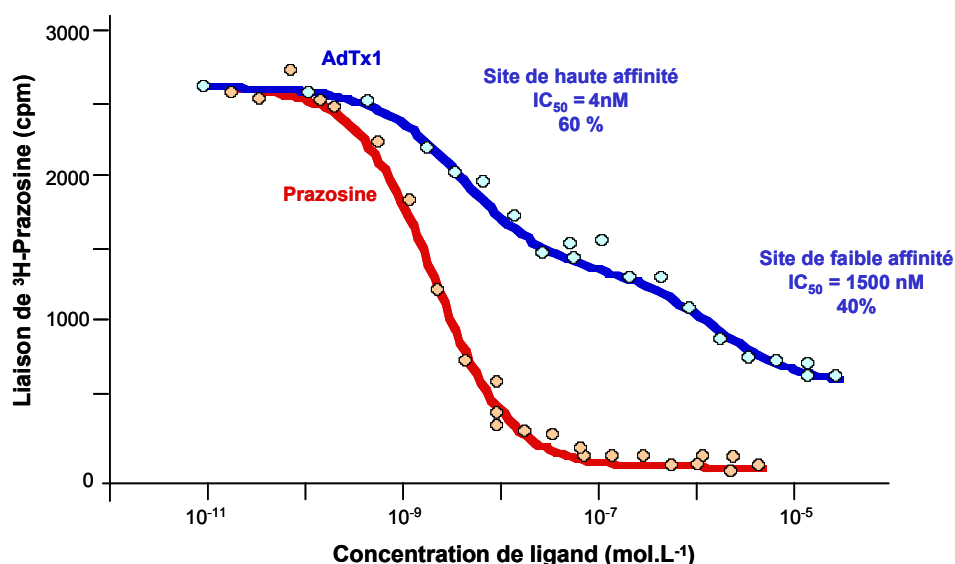


Figure 20: Déplacement de la liaison de ^3H -prazosine avec les récepteurs adrénrgiques de synaptosomes de rat par la prazosine non tritiée (en rouge) ou l'AdTx1 (en bleu)

La courbe rouge présente le déplacement de la liaison de la prazosine tritiée par la prazosine non marquée. L' IC_{50} de la courbe (point d'inflexion) est de l'ordre de $1,1 \text{ nmol.L}^{-1}$. Sur la courbe bleue, qui indique le déplacement de la liaison par l'AdTx1, on observe deux points d'inflexion, ce qui signifie que deux sites de liaisons de la prazosine tritiée sont perturbés par l'AdTx1. La prazosine est un ligand spécifique des récepteurs de type α_1 -adrénrgiques mais elle lie indifféremment les trois sous-types α_{1a} , α_{1b} et α_{1c} . La *Figure 20* illustre la limitation des expériences ligand-récepteur qui ne permettent pas de déterminer quels sous-types sont mis en jeu.

Pour répondre à cette question, deux types d'expériences sont envisageables. Le premier consiste à déplacer la liaison d'un ligand spécifique d'un sous-type en particulier. Malheureusement, ceci n'est pas possible pour nos études car aucun ligand spécifique d'un de ces sous-types est disponible. La deuxième possibilité est de travailler avec des membranes qui n'expriment qu'un seul sous-type de récepteur avec la prazosine comme ligand. Il est alors plus aisé de déterminer clairement l'affinité de la toxine pour le sous-type de récepteur exprimé. C'est donc cette seconde option que nous avons choisie. Des membranes de levures n'exprimant que des sous-récepteurs α_{1A} , α_{1B} ou α_{1D} ont été utilisées pour évaluer l'affinité de l'AdTx1 pour ces derniers (*Figure 21*). A titre de comparaison, l'AdTx1 a également été testée par la même méthode sur deux autres types de récepteurs adrénrgiques, l' α_{2c} et le β_2 .

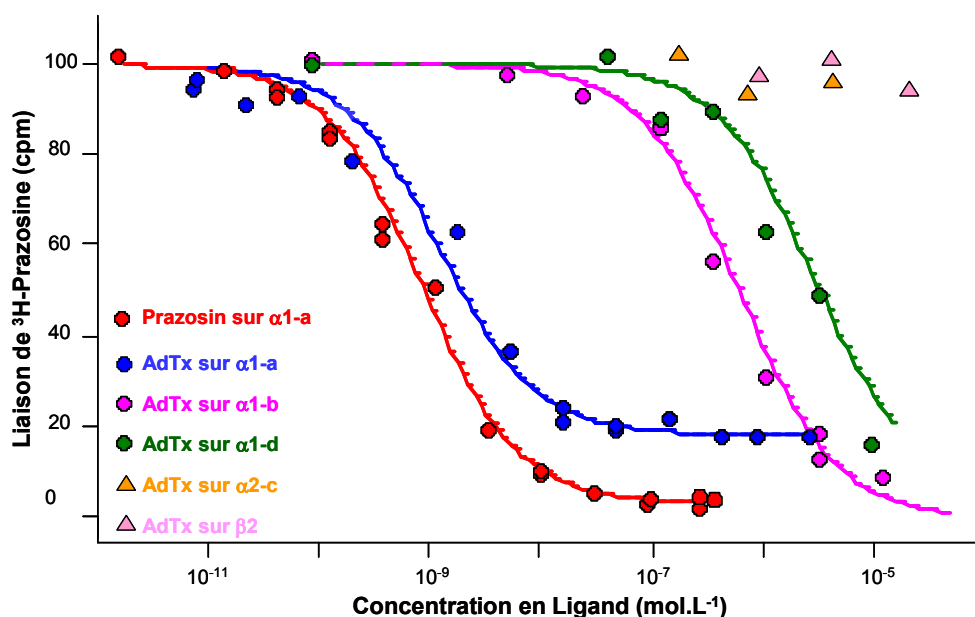


Figure 21: Déplacement de la liaison de ³H-prazosine par la prazosine froide et par l'AdTx1 sur les sous-types de récepteurs α_{1A} , α_{1B} , α_{1C} mais également α_{2C} et β_2 pour comparaison

Ce graphique montre clairement que l'AdTx1 possède une bien meilleure affinité (IC₅₀ plus faible) pour les sous-récepteurs α_{1A} (courbe bleue) que pour les α_{1B} (courbe violette) et pour les α_{1D} (courbe verte). Cela se traduit par une sélectivité de la toxine pour les récepteurs α_{1A} respectivement 1000 et 2000 fois supérieure à celles mesurées pour les α_{1B} et les α_{1D} . La Figure 21 montre également que l'AdTx1 ne se lie pas avec une bonne affinité aux deux autres sous-types de récepteurs adrénergiques testés α_{2C} et β_2 .

La sélectivité marquée de l'AdTx1 pour les sous-types α_{1A} peut être exploitée à plusieurs niveaux :

- D'un point de vue pharmacologique (voire thérapeutique), l'AdTx1 permettrait d'agir directement sur le sous-type α_{1A} . Ce récepteur est impliqué dans le contrôle de la tonicité de certains muscles génitaux, de la vessie et de la prostate. Elle constitue donc un bon candidat pour le développement de médicaments liés à des pathologies impliquant ces muscles. A titre d'exemple, l'hypertrophie bénigne de la prostate chez les sexagénaires, qui entraîne des problèmes gênants d'incontinence urinaire, implique des muscles prostatiques, cibles potentielles de l'AdTx1.

- L'AdTx1 permet de cibler sélectivement les récepteurs α_{1A} . Il est clair que pour étudier leur structure et leur distribution au sein des différentes cellules, un outil comme l'AdTx1 est très intéressant. Pour rendre cet outil opérationnel, il est primordial de le rendre visible, c'est-à-dire de le marquer chimiquement (radioactivité, fluorescence) afin de pouvoir

suivre sont évolution et sa distribution dans les cibles testées. Le développement d'un outil pharmacologique fluorescent et sa caractérisation par spectrométrie de masse FT-ICR font l'objet de la partie suivante.

V. Caractérisation structurale d'un homologue fluorescent de l'AdTx1 par FT-ICR

A. Contexte général

La sélectivité de l'AdTx1 pour les récepteurs α_{1A} adrénergiques est très intéressante puisque aucun anticorps spécifique n'est disponible et qu'aucun ligand marqué n'existe.

Une modification chimique de l'AdTx1 pour en faire un ligand fluorescent spécifique de α_{1A} serait donc extrêmement utile. Le fait que ce ligand soit un peptide permet plusieurs types de marquages dont les plus courants sont le marquage radioactif et le marquage fluorescent.

- Le marquage radioactif

Les deux isotopes radioactifs couramment utilisés pour la fabrication de ligands radiomarqués sont l'iode 125 (^{125}I) et le tritium (^3H). Dans le cas de l'iode, deux voies majeures d'incorporation de ^{125}I sur le peptide existent. La première consiste à greffer par substitution nucléophile un atome d' ^{125}I directement sur le groupement aromatique d'une tyrosine ou de la chaîne latérale d'une histidine. La seconde technique consiste à greffer sur le peptide un réactif chimique contenant déjà un ^{125}I (réactif de Bolton et Hunter modifiant les lysines et les extrémités N-terminales). Cette technique est principalement utilisée lorsque le peptide ne contient pas de tyrosine permettant l'iodation directe ou quand l'activité du peptide dépend fortement de l'intégrité de l'acide aminé réactif.

Plusieurs inconvénients du marquage à l'iode sont à noter. Tout d'abord, le temps de demi-vie du radioélément ^{125}I étant assez court (60 jours), l'utilisation du ligand marqué dans les 60 jours qui suivent sa synthèse est nécessaire. Ensuite, lorsque le marquage se situe près du site actif, l'atome d'iode, volumineux et hydrophobe peut affecter l'affinité du ligand pour son récepteur. Enfin, ce type de marquage ne peut s'effectuer que sur certains types d'acides aminés ou sur certaines positions dans le squelette polypeptidique (K, Y, H et extrémité N-

terminale). Pour éviter ces différents inconvénients, il est possible de marquer les peptides avec du tritium (^3H). Les deux modes d'incorporation décrits pour l'iode sont utilisables. Le marquage direct peut être effectué par échange isotopique avec le ligand (échanges $\text{H}/^3\text{H}$) tandis que le marquage par l'intermédiaire d'une molécule organique marquée nécessite également la présence de lysines ou d'une extrémité *N*-terminale accessible. L'inconvénient du marquage direct est que le ligand radioactif doit être conservé dans un solvant aprotique pour éviter les phénomènes de rétro-échanges isotopiques. L'avantage du marquage au tritium est, qu'une fois marqués, les ligands peuvent être utilisés longtemps puisque la demi-vie du tritium est de 12,3 ans. Le remplacement d'un ^1H par un ^3H n'induit pas de modification de structure majeure et ne modifie donc pas les capacités de liaison du ligand. Un autre avantage est que l'on peut marquer tous les peptides, ce qui n'est pas le cas des marquages à l'iode.

Le problème lié au marquage radioactif est que la manipulation des ligands radiomarqués nécessite des précautions et des équipements particuliers et très coûteux. Le stockage de ce type de composés est également très réglementé et impose de nombreuses contraintes. Dans ce contexte, le marquage par fluorescence offre une alternative intéressante.

- Le marquage fluorescent

Plusieurs types de fluorophores peuvent être greffés directement sur les peptides dont la détection peut alors s'effectuer par microscopie confocale. Les principaux fluorophores disponibles commercialement réagissent sur les amines réactives présentes dans les peptides. Néanmoins, comme dans le cas de l'iode, l'ajout d'un fluorophore peut avoir des conséquences sur l'activité du peptide modifié. La taille et la nature du fluorophore vont influencer directement sur l'encombrement stérique et l'hydrophobicité de la molécule se traduisant notamment par une modification de l'affinité du ligand pour son récepteur.

La microscopie confocale permet une très bonne résolution (cellules en cultures, coupes fines d'organes ...) ce qui en fait une technique très utilisée pour l'étude de la liaison de ligands à leurs récepteurs^{7,8}.

B. Caractérisation du mutant AdTx1 (K₃₄R*) par une approche « top-down » par ECD

En tenant compte de tous ces éléments, l'AdTx1 a été marquée par un fluorophore dérivé de la cyanine appelé Cy3BTM⁹ qui sera présenté en détail un peu plus loin. Ce composé organique réagit sur les amines des lysines et à un degré moindre avec l'amine primaire de l'extrémité N-terminale. La séquence de l'AdTx1 possède six sites de modification potentiels (5 lysines et l'extrémité N-terminale). L'un d'entre eux est situé en bas de la boucle 2 de la toxine à trois doigts (K₃₄). Il est connu que cette charge positive constitue le point d'ancrage de toutes les toxines à trois doigts à leur récepteur. Elle doit donc être maintenue impérativement pour ne pas affecter la liaison avec son récepteur. Le problème est que cette lysine est susceptible d'être modifiée par le fluorophore Cy3BTM.

Pour éviter cette modification, un mutant a été synthétisé dans lequel la lysine K₃₄ a été remplacée par une homoarginine R₃₄* (arginine dont la chaîne latérale est rallongée d'un groupement méthylène). Ce site devient non réactif au fluorophore tout en maintenant la charge positive en place, primordiale pour la liaison. Des tests de liaison ont montré que la différence de charge qui existe entre la lysine et l'homoarginine ne modifie pas significativement la liaison de l'AdTx1 à son récepteur. Les toxines qui possèdent une lysine en position +34 sont minoritaires au sein des toxines à trois doigts puisque 90% d'entre elles présentent une arginine à cette position. La mutation de la lysine en *homo*arginine (R*) (et pas simplement en arginine) avait également pour but d'évaluer l'influence de la longueur de la chaîne latérale du résidu sur la liaison avec le récepteur. Les résultats biologiques obtenus grâce à cette étude ne seront pas abordés dans cette thèse.

L'étape de mutation (K₃₄ → R₃₄*) est cruciale puisqu'elle prépare la toxine à l'étape suivante de modification par le fluorophore. Aussi, une caractérisation du mutant par spectrométrie de masse FT-ICR a été effectuée. Etant donné les résultats encourageants obtenus par l'approche « top-down » sur la toxine native, la caractérisation de la forme mutée a été réalisée par ECD dans le but de localiser directement la modification par l'analyse des fragments obtenus en phase gazeuse.

Préalablement à cette analyse, les ponts S-S de la toxine mutée ont été réduits par le TCEP. La toxine réduite, purifiée sur Ziptip[®] C18 et éluée par le solvant de nano-electrospray, permet d'obtenir le spectre nanoESI-FT-ICR reporté sur la *Figure 22*.

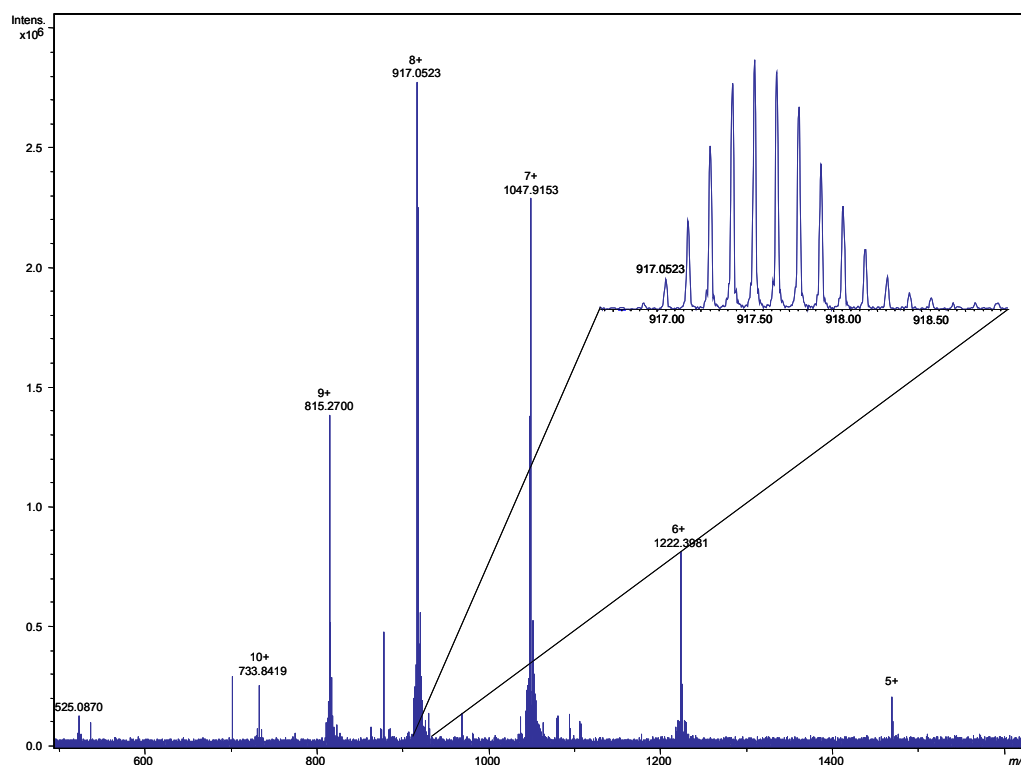


Figure 22 : Spectre nanoESI-FT-ICR de l'AdTx1 mutée

L'analyse des résultats indique la présence très majoritaire d'une toxine. Sa masse mesurée est 7328,3602 Da, ce qui fait une différence de +42,00 Da avec la masse de la toxine naturelle. Ceci correspond exactement à la différence de masse attendue entre une lysine (128,0949 Da) et une homoarginine (170,1168 Da). La masse théorique de la toxine mutée étant de 7328,3837 Da, l'erreur sur la mesure de masse est donc de 3,2 ppm, ce qui est tout à fait correct.

La fragmentation de cette espèce majoritaire par ECD conduit à un grand nombre de fragments (*Figure 23*) résultant de la rupture des liaisons C_{α} -N (ions *c* et *z*). Une nouvelle fois, des ions *y* sont observés et permettent d'apporter des informations structurales complémentaires.

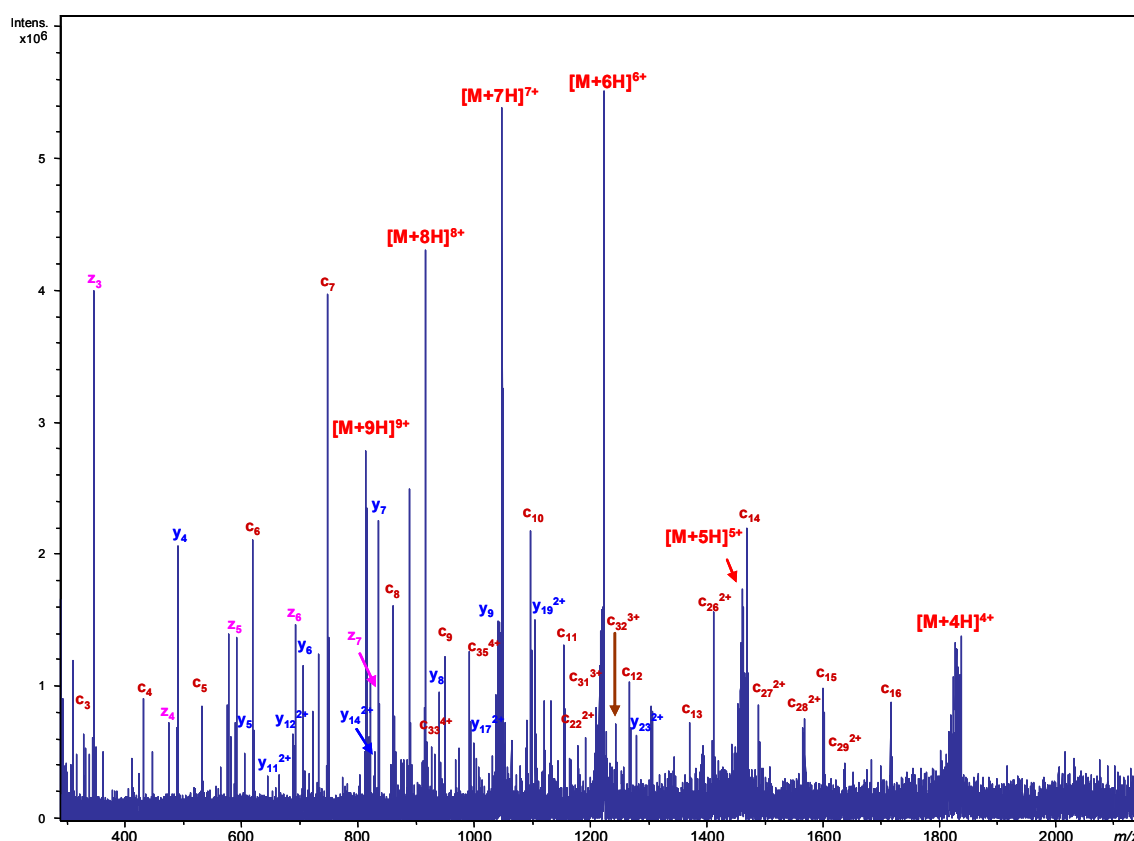


Figure 23: Spectre ECD de la toxine mutée



Les séries d'ions c et y permettent d'obtenir quasiment toute la séquence de la toxine. De ce fait, la position de la mutation de la lysine K_{34} en homoarginine R^*_{34} peut facilement être mise en évidence par la différence de masse entre les ions c_{33} et c_{35} qui est de 283,2309 Da pour une valeur théorique de 283,2008 Da. Le mutant présente donc bien la séquence attendue.

C. Transformation de l'AdTx1 en ligand fluorescent

C.1. Contexte général et réactivité du Cy3B™

La fluorescence est une méthode de détection extrêmement répandue pour les études pharmacologiques. La limite de détection de cette technique est proche de celle des

marquages radioactifs et permet de s'affranchir totalement des inconvénients liés à la manipulation de radioéléments. Des composés dérivant des cyanines sont utilisés depuis plus de dix ans comme sondes de fluorescence¹⁰. La cyanine utilisée pour le marquage de l'AdTx1 est le Cy3B™ (*Figure 24*) d'Amersham Biosciences (Cardiff, United Kingdom).

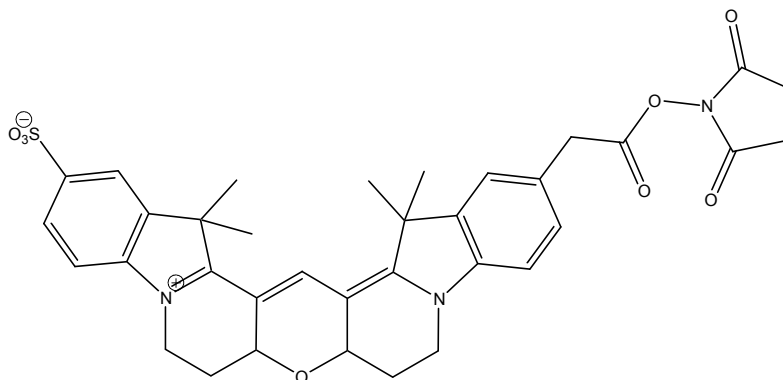


Figure 24: Structure du fluorophore Cy3B™

Ce réactif est spécifique des amines primaires. Sur l'AdTx1(R*), cinq sites sont accessibles après la protection de la lysine 34 : 4 lysines et le groupement N-terminal.

Une purification du mélange réactionnel est effectuée par chromatographie en phase inverse. Quatre pics chromatographiques sont obtenus caractérisant la présence des formes non modifiées et marquées une, deux ou trois fois. Chacune des fractions est analysée par spectrométrie de masse et son activité biologique sur les récepteurs d'intérêt est évaluée. Les résultats montrent qu'une baisse d'affinité d'un facteur 10^n (avec n = nombre de fluorophores greffé) est observée pour les toxines modifiées. Ainsi, la fraction représentant le meilleur compromis [toxine fluorescente et perte limitée (un facteur 10 est acceptable)] est celle contenant la toxine marquée par un seul fluorophore. Le spectre nanoESI-FT-ICR obtenu pour cette fraction est reporté sur la *Figure 25*.

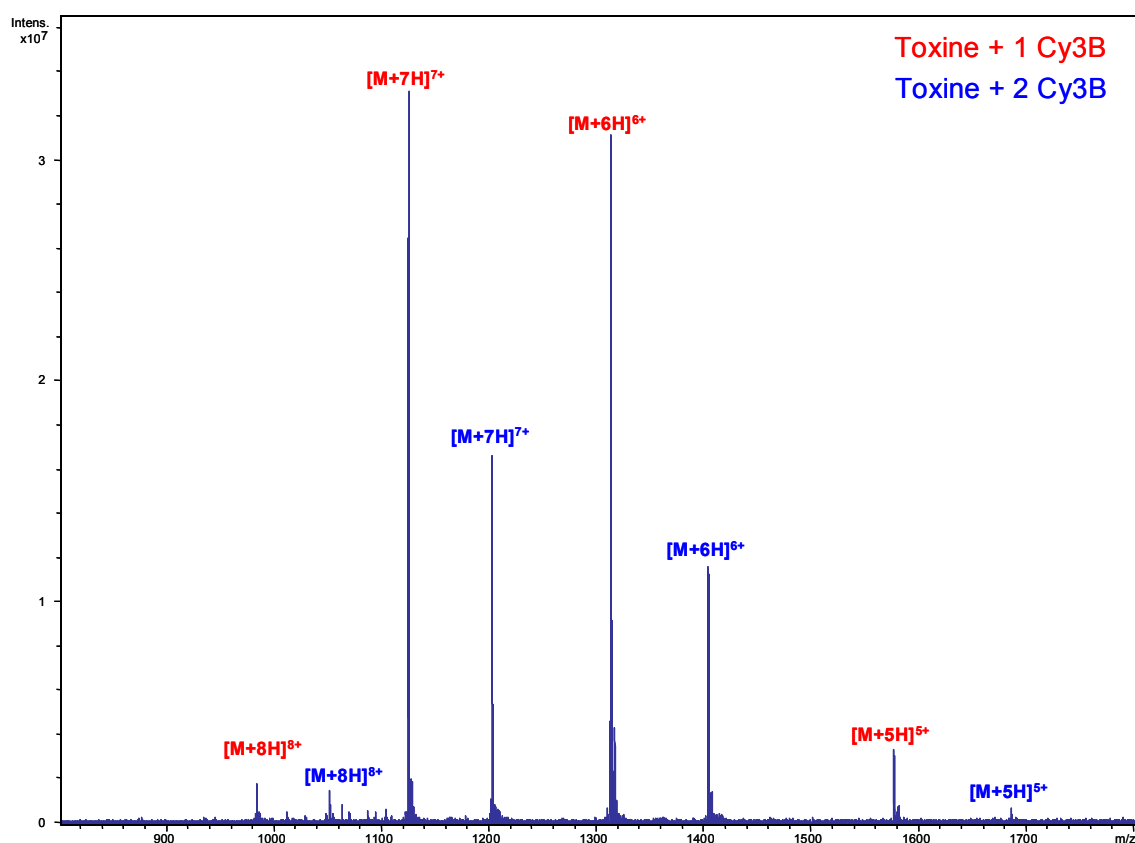


Figure 25: Spectre nanoESI-FT-ICR de l'AdTx1-R* fluorescente. En rouge : toxine modifiée par un fluorophore. En bleu : toxine marquée par deux fluorophores

L'analyse de ce spectre révèle la présence de deux espèces dont les masses correspondent parfaitement aux masses théoriques de l'AdTxR* modifiée par un et deux fluorophores Cy3B. Ce mélange était attendu puisque la séparation chromatographique du milieu réactionnel ne permet pas l'obtention de pics bien résolus. Cette séparation médiocre s'explique par une grande homologie, au niveau structural, entre les différentes toxines modifiées qui conduit à des espèces ayant un comportement hydrophobe assez similaire. Ainsi, la fraction collectée sensée correspondre à la toxine modifiée une seule fois contient également une proportion non négligeable de toxine modifiée deux fois.

Espèce	AdTxR* + 1Cy3B	AdTxR* + 2 Cy3B
Masse mesurée (Da)	7870,6370	8412,8286
Masse théorique (Da)	7870,5712	8412,7578
ΔM (ppm)	8,2	8,4

Cette expérience permet de vérifier que la réaction chimique s'est déroulée convenablement et de dénombrer le nombre de fluorophores greffés. Une analyse « top-down » a alors été réalisée pour localiser les fluorophores dans la séquence. La localisation des sites modifiés n'a pas de réel intérêt sur le plan biologique puisque la seule lysine importante a été modifiée en homoarginine. En revanche, cela représente un intérêt du point de vue du spectrométriste de masse non seulement sur le plan méthodologique mais aussi afin de poursuivre notre évaluation de l'approche « top-down ».

C.2. Localisation des sites modifiés par une approche « top-down » par dissociation en source (ISD)

La première étape est l'isolation des ions $[M+6H]^{6+}$ correspondant à la toxine modifiée une fois puisque la fraction n'est pas pure. Une fois cette sélection réalisée (par éjection de tous les autres ions), les parents $[M+6H]^{6+}$ ont été irradiés par le faisceau d'électrons de faible énergie. Malheureusement, le résultat est assez décevant. En effet, très peu de fragments sont observés et aucune information sur la position des modifications n'est obtenue. La raison principale pour expliquer ce résultat est sans doute la faible intensité du signal correspondant aux ions précurseurs $[M+6H]^{6+}$ après l'étape d'isolation. Il est également possible que la présence du fluorophore ait une influence sur l'efficacité de la capture de l'électron.

Des expériences d'ECD, directement sur le mélange des deux formes, ont alors été réalisées. Notre objectif était d'obtenir un signal initial plus intense afin d'obtenir un spectre ECD plus informatif (même si la relation entre un ion fragment et un nombre de modification précis est plus délicate du fait de l'absence de sélection d'ions parents).

Le spectre obtenu contient notamment une belle série d'ions *c* ainsi que quelques ions *b* (*Figure 26*), mettant en évidence l'importance de l'intensité initiale du signal. Le spectre est néanmoins moins informatif que celui obtenu lors de la caractérisation du mutant et les ions observés ne permettent pas de conclure définitivement sur la position des modifications. Deux informations majeures peuvent être cependant tirées de ce spectre et concernent la lysine K₇ et l'extrémité N-terminale.

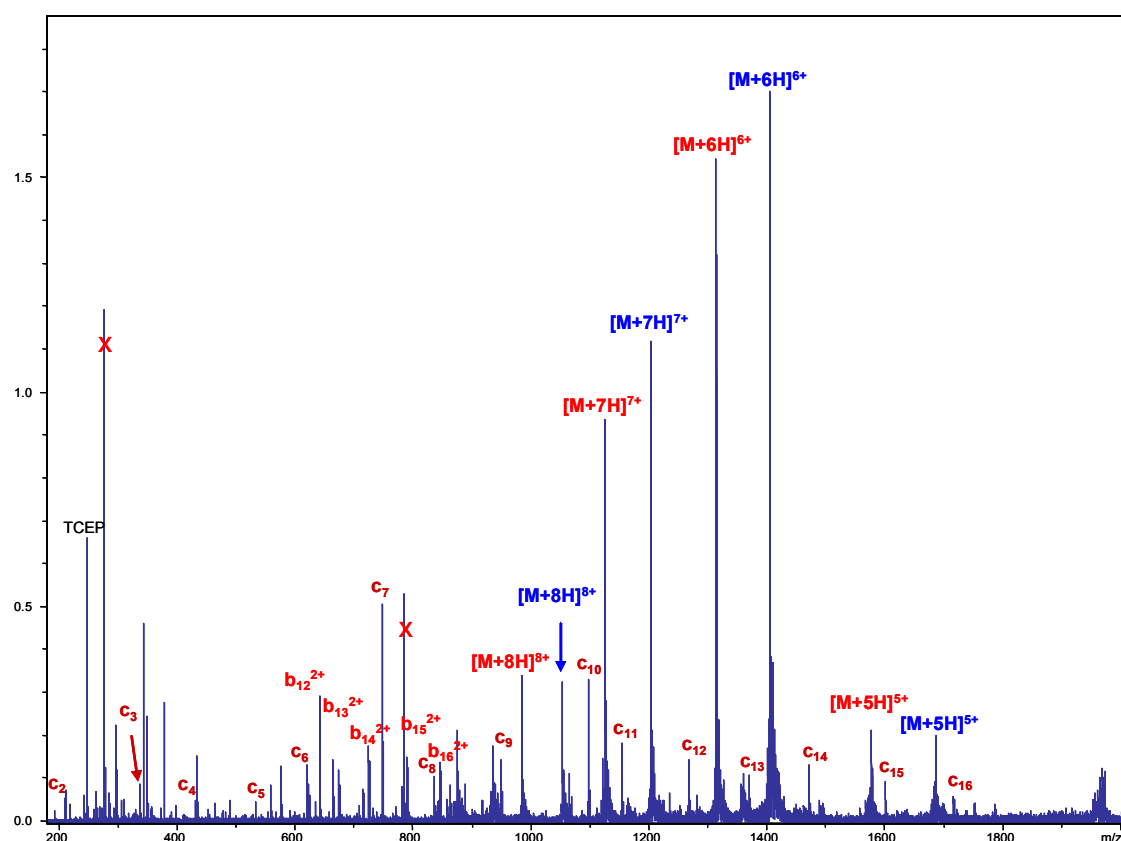
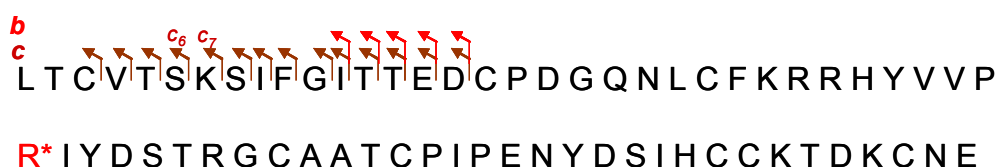


Figure 26: Spectre ECD obtenu pour les deux formes de l'AdTx1(R*) fluorescente

La masse exacte d'une lysine non modifiée est retrouvée entre les ions c_6 et c_7 annotés sur le spectre, caractérisant le fait que cette lysine n'a pas été affectée par la réaction avec le Cy3B™. Aucune trace de sa forme modifiée n'est détectée. De la même façon, on peut mettre en évidence que l'amine primaire de l'extrémité *N*-terminale n'est pas non modifiée car tous les ions observés dans le spectre ont des masses identiques à celles calculées pour une extrémité libre. Ces résultats permettent de supposer qu'aucune des deux formes modifiées de l'AdTx1(R*) ne porte un fluorophore en *N*-terminal et sur la lysine +7.



Afin de compléter ces résultats obtenus sur protéines entières, nous avons décidé de fragmenter les toxines directement en source par ISD (*In Source Decay*) par l'application d'un haut potentiel (200 à 400 V) sur la contre électrode. La difficulté est, une nouvelle fois, qu'il n'y a pas de possibilité de sélection d'une espèce particulière puisque la fragmentation se

située avant l'analyseur. Les deux toxines fluorescentes seront donc soumises à cette fragmentation.

La première expérience ISD a été effectuée en appliquant un potentiel de 300 V sur la lentille d'extraction (Figure 27).

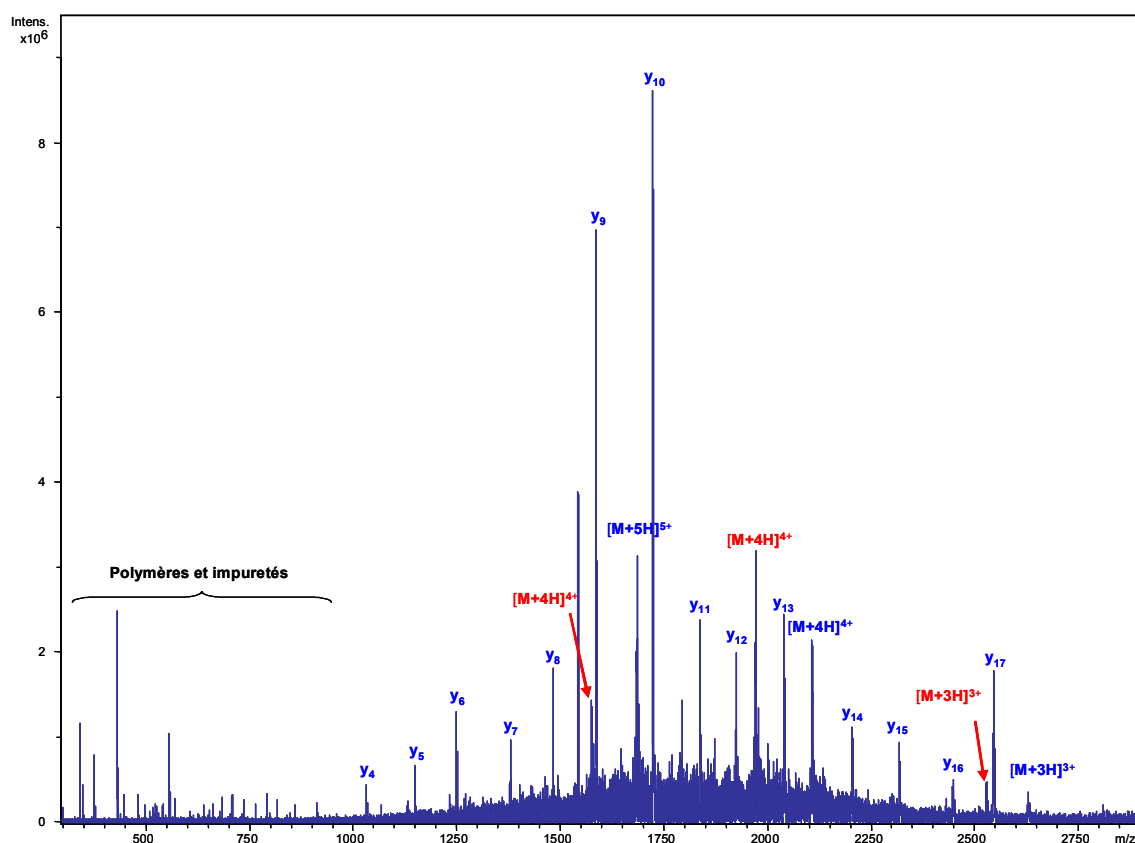
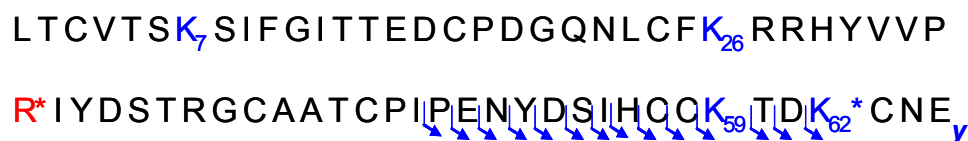


Figure 27: Spectre ISD obtenu en nanoESI-FT-ICR à partir du mélange d'AdTx(R*) fluorescentes - Potentiel appliqué -300 V - 50 accumulations

Une belle série d'ions y est obtenue. Elle permet de mettre en évidence de manière non ambiguë que la lysine 59 n'est pas modifiée : (i) la différence entre les ions y_6 et y_7 nous permet de mesurer exactement la masse d'une lysine (ii) aucun signal correspondant à ce même acide aminé modifié n'est détecté. Il est à noter qu'aucune perte du Cy3BTM, à partir des ions parents mais également à partir des différents ions y , n'est observée. Ce point est important pour l'attribution des pics dans le spectre. La présence d'un ion y_4 à m/z 1035,4847 est intéressante puisqu'elle permet de déterminer que la lysine +62 est modifiée par un Cy3B (aucun ion y_4 sans Cy3BTM n'est observé confirmant que la liaison NH-CO avec le fluorophore ne se fragmente pas lors des collisions). Les rapports m/z attendus pour l'ion y sont : m/z 493,2070 s'il ne possède pas de modification et m/z 1035,3945 avec un groupement

Cy3B. L'extrémité C-terminale n'étant pas réactive, il apparaît clairement que la lysine +62 porte un fluorophore.



Cependant, il est impossible de tirer d'autres conclusions de ce spectre. A ce stade, nous avons mis en évidence par nos différentes approches sur protéines entières que les lysines K_7 et K_{59} ne semblent pas présenter de modification alors que la lysine K_{62} est marquée par un fluorophore. Aucune information sur la lysine K_{26} n'est disponible. Par élimination, et étant donnée qu'une espèce modifiée deux fois est présente dans le spectre, il semble probable que cette dernière soit modifiée. Des nouvelles expériences en ISD avec un potentiel d'accélération plus important ont donc été menées (Figure 28).

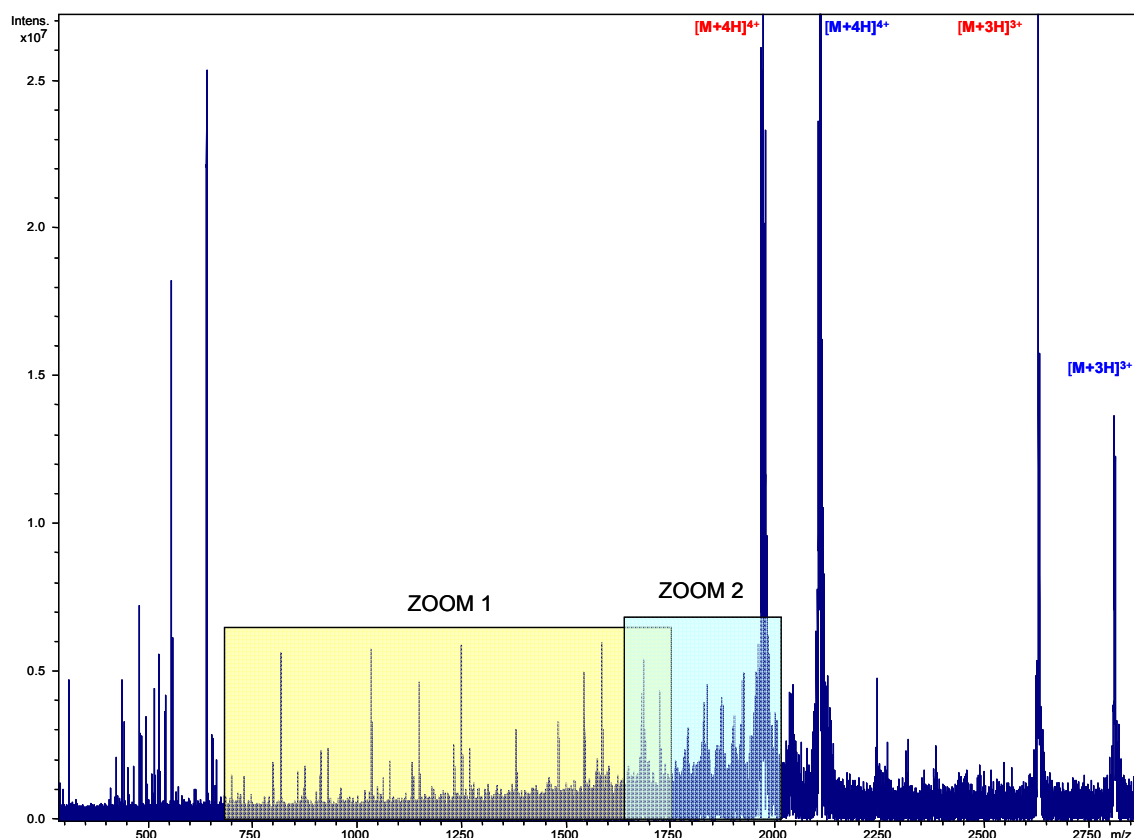


Figure 28: Spectre ISD obtenu en nanoESI-FT-ICR des AdTx(R^*) fluorescentes - Potentiel appliqué -400 V - 130 accumulations

Pour une lecture plus facile, deux agrandissements de ce spectre sont proposés sur la Figure 29.

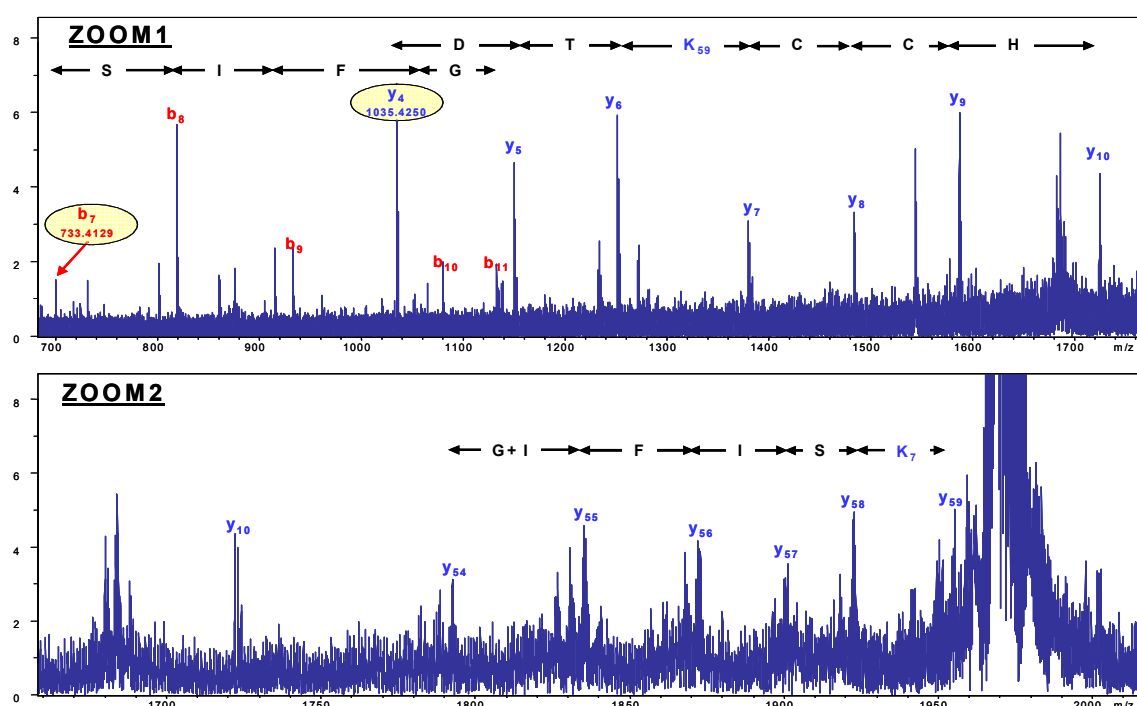
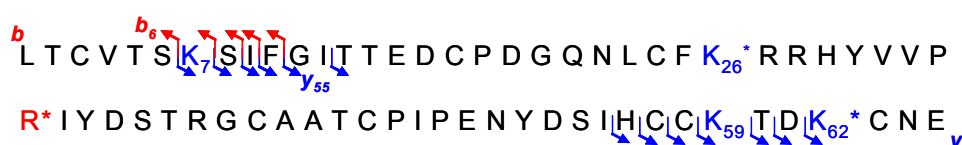


Figure 29 : Agrandissements de la Figure 38 - (A) m/z [700 - 1800] (B) m/z [1750 - 2020]

Le premier agrandissement confirme nos premières observations, à savoir que l'extrémité *N*-terminale et les lysines K_7 et K_{59} ne portent pas de modification alors que la lysine +62 est modifiée par un Cy3B. Le second agrandissement nous permet d'observer des ions caractéristiques d'une autre région de la séquence de l'AdTx1(R*) fluorescente. Ainsi, la masse de l'ion y_{55} , déconvoluée à 7480,4449 Da, coïncide parfaitement avec la masse du peptide correspondant modifié par deux Cy3B™ calculée à 7480,2587 Da contre 6938,0711 Da pour une seule modification et 6395,8836 Da sans modification. Les lysines marquées d'un astérisque sont celles qui ont été trouvées modifiées. Aucune donnée spectrale ne permet de mettre en évidence la présence de plusieurs formes.



Il apparaît donc que les lysines K₂₆ et K₆₂ portent majoritairement les deux fluorophores. Néanmoins, rien ne permet d'affirmer que d'autres formes n'existent pas.

Pour lever cette ambiguïté, une digestion à la trypsine de la toxine fluorescente est effectuée. Le mélange de peptides obtenu est ensuite analysé par nanoESI-Q-TOF. La carte peptidique révèle un mélange de différentes formes puisque des peptides modifiés non seulement sur les lysines K₂₆ et K₆₂ sont retrouvés mais également sur la lysine K₇ qui n'avait pas été mis en évidence lors de l'étude « Top-Down ». En revanche, cette analyse ne permet pas de mettre en évidence que la K₅₉ est partiellement modifiée puisque aucun peptide caractéristique n'a pu être détecté. Le spectre montre également que les peptides les plus intenses correspondent aux modifications en position K₂₆ et K₆₂, la lysine K₇ modifiée n'étant représentée que par un seul peptide de très faible intensité. Nos expériences « bottom-up » sont en total accord avec les hypothèses émises lors des études « top-down », ce qui met en évidence une nouvelle fois la complémentarité de ces deux approches.

VI. Conclusion

Nous avons, au cours de ce travail, appliqué la spectrométrie de masse FT-ICR à l'étude d'une toxine de haut poids moléculaire (7,3 kDa) issue du venin du mamba *Dendroaspis angusticeps* et de plusieurs de ses homologues portant un fluorophore. Bien que ce venin ait déjà été étudié, cette toxine n'a jamais été décrite auparavant. D'un point de vue biologique, elle est active sur les récepteurs α_{1A} adrénergiques dont aucun ligand spécifique n'était connu jusqu'alors. Sa séquence, comportant 65 acides aminés, est très proche de celles des toxines muscariniques, actives sur les récepteurs muscariniques, une famille de RCPG (Récepteurs Couplés à la Protéine G).

Les images obtenues par le Dr Jordi Molgo (CNRS, Gif-sur-Yvette) par microscopie confocale sont présentées *Figure 30*.. La figure (A) montre le résultat obtenu après que des cellules modèles surexprimant le récepteur α_{1A} -adrénergiques aient été incubées en présence d'Adtx1 fluorescente. L'image montre clairement que la toxine lie sélectivement les récepteurs α_{1A} présents au niveau de la paroi cellulaire. La figure (B) montre que le réactif fluorescent Cy3BTM seul ne présente aucune affinité pour les cellules modèles, mettant ainsi en valeur l'affinité spécifique de la toxine pour les récepteurs surexprimés.

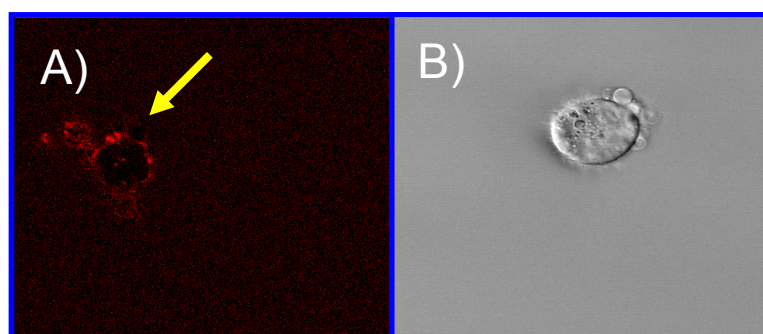


Figure 30: Localisation des récepteurs α_{1A} -adrénergiques de cellules modèles par l'AdTx1 fluorescente sur des cellules surexprimant – (A) Incubation des cellules avec 1 μ M d'AdTx1 fluorescente – (B) Incubation des cellules avec le réactif fluorescent Cy3B

L'originalité du travail de spectrométrie de masse réside dans le fait que nous avons testé deux approches différentes pour caractériser cette toxine : (i) l'approche classique « bottom-up » par digestion enzymatique et fragmentation des peptides obtenus (ii) l'approche « top-down » sur protéine entière.

L'analyse « bottom-up » s'est révélée très efficace puisque nous avons pu déterminer les 15 derniers acides aminés contenant l'extrémité C-terminale de la toxine étudiée, qui manquaient après la dégradation d'Edman. Nous montrons, encore une fois, que la précision apportée par la spectrométrie de masse FT-ICR est un élément-clé pour le séquençage *de novo*. L'approche « top-down » par ECD (fragmentation de la toxine native mais également modifiée par des fluorophores en phase gazeuse) est une alternative intéressante car elle permet en principe d'obtenir, en une seule expérience, des informations de structure primaire sur l'ensemble de la séquence. Néanmoins, nos expériences montrent aussi que les fragments observés en ECD proviennent majoritairement des extrémités N- et C-terminales des peptides et que la partie centrale, moins accessible, se fragmente mal. Ce phénomène s'explique simplement par la structuration de la protéine en phase gazeuse, qui conduit à la formation de liaisons faibles (liaisons hydrogène par exemple), qui ne sont pas rompues lors des expériences ECD. Une solution, pour pallier ce problème, consiste à activer la protéine en phase gazeuse avant l'ECD, pour rompre le réseau de liaisons faibles. Cette activation est généralement réalisée par des photons infra-rouge de faible énergie. L'implémentation d'un laser à CO₂ permettant de réaliser ces expériences IRMPD/ECD sur l'instrument du laboratoire est en cours.

Une autre limitation de l'approche « top-down » est qu'il est nécessaire de partir d'un signal intense dans le spectre avant ECD car cette technique de dissociation est peu efficace. Nous montrons dans ce travail que l'étape d'isolation d'une espèce donnée, pour la fragmenter

sélectivement, peut conduire à une perte de signal importante et à des spectres ECD peu informatifs. Ce problème est moins crucial sur les instruments FT-ICR de nouvelle génération (hybrides), car l'ion parent d'intérêt peut être concentré (par sélection dans un autre analyseur) avant son entrée dans la cellule.

Références Bibliographiques

-
- ¹ Perkins J.R., Parker C.E., Tomer K.B., The characterization of snake venoms using capillary electrophoresis in conjunction with electrospray mass spectrometry: Black mambas, **1993**, *Electrophoresis*, 14, 458-468.
- ² Max S.I., Liang J.S., Potter L.T., Purification and properties of m1-toxin, a specific antagonist of m1 muscarinic receptors, **1993**, *J. Neurosci.*, 13, 4293-4300.
- ³ Carsi, J.M., Potter L.T., m1-Toxin isotoxins from the green mamba (*Dendroaspis angusticeps*) that selectively block m1 muscarinic receptors, **2000**, *Toxicon*, 38, 187-198.
- ⁴ Kukhtina V.V., Weise C., Muranova T.A., Starkov V.G., Franke P., Hucho F., Wnendt S, Gillen C, Tsetlin V.I., Utkin Y.N., Muscarinic toxin-like proteins from cobra venom, **2000**, *Eur. J. Biochem.*, 267, 2784-2789.
- ⁵ Miwa J.M., Ibanez-Tallon I., Crabtree G.W., Sanchez R., Sali A., Role L.W., Heintz N., Lynx1, an endogenous toxin-like modulator of nicotinic acetylcholine receptors in the mammalian CNS, **1999**, *Neuroscience*, 23, 105-114.
- ⁶ Lawson R., Slowinski J.B., Crother B.I., Burbrink F.T., Phylogeny of the *Colubroidea* (Serpentes): New evidence from mitochondrial and nuclear genes, **2005**, *Mol Phylogent. Evol.*, 37, 581-601.
- ⁷ Massensini A.R., Suckling J., Brammer M.J., Moraes-Santos T., Gomez M.V., Romano-Silva M.A., Tracking sodium channels in live cells: confocal imaging using fluorescently labelled toxins, **2002**, *J Neurosci Methods*, 116, 189-196.
- ⁸ Barbier J., Lamthanh H., Le Gall F., Favreau P., Benoit E., Chen H., Gilles N., Ilan N., Heinemann S.H., Gordon D., Menez A., Molgo J., A δ -conotoxin from *Conus ermineus* venom inhibits inactivation in vertebrate neuronal Na⁺ channels but not in skeletal and cardiac muscles, **2004**, *J. Biol. Chem.*, 279, 4680-4685.
- ⁹ Cooper M., Ebner A., Briggs M., Burrows M., Gardner N., Richardson R., West R., Cy3B, Improving the performance of cyanine dyes, **2004**, *J. Fluoresc.*, 14, 145-150.
- ¹⁰ Mujumdar R.B., Ernst L.A., Mujumdar S.R., Lewis C.J., Wagonner A.S., Cyanine dye labelling reagent : sulfoindocyanine succinimidyl ester, **1992**, *Histochem.*, 98, 81-85.

Conclusion

Nous avons, au cours de ce travail de thèse, utilisé la spectrométrie de masse FT-ICR pour explorer différents axes de recherche visant à déterminer la séquence de nouvelles toxines dans des venins bruts. Nous nous sommes intéressés à deux familles très différentes de venins : les venins de cônes marins, contenant majoritairement des petits peptides très structurés et modifiés post-traductionnellement et des venins de serpent, présentant souvent des composés de plus haut poids moléculaire. Dans tous les cas, nous avons cherché à exploiter les performances particulières de la spectrométrie de masse FT-ICR (précision, résolution et méthodologie ECD) pour mettre au point des stratégies d'analyse innovantes.

Dans ce contexte, nous avons montré pour la première fois qu'il est tout à fait possible de réaliser le séquençage *de novo* (dans le sens réel du terme puisque les organismes sur lesquels nous travaillons ne sont pas séquencés) de peptides se trouvant dans les venins bruts, sans séparation préalable. Une étude combinée de séquençage *de novo* et de clonage moléculaire a également montré tout l'intérêt d'établir des cartographies massiques de venin avec une grande précision sur la mesure de masse. Nos résultats indiquent néanmoins que cette approche, sur venin brut, ne peut plus être appliquée lorsque le venin est extrêmement complexe et qu'il faut alors procéder à son fractionnement (par chromatographie mono ou bidimensionnelle) avant analyse par spectrométrie de masse. Une comparaison, pour le même venin, des approches en infusion directe par nanoelectrospray et en couplage par nanoLC-FT-ICR liquide montre sans ambiguïté que le couplage permet d'améliorer la gamme dynamique et que les deux approches sont complémentaires. En effet, si l'utilisation du couplage permet de déterminer un plus grand nombre de masses moléculaires dans le venin, on observe la perte de certains composés, pourtant observés en infusion directe.

Les résultats les plus intéressants du point de vue méthodologique sont ceux obtenus pour le développement d'une approche « top-down » en utilisant la technique de Dissociation par Capture d'Electron. Cette approche repose sur l'identification de la protéine d'intérêt par sa fragmentation directe en phase gazeuse (par opposition à l'approche « bottom-up » dont la première étape est la digestion de la protéine à identifier). Nous avons ainsi caractérisé une toxine de 8 kDa environ et plusieurs de ses homologues portant des fluorophores. La rapidité, et la facilité de mise en oeuvre de cette technique (surtout lorsque l'espèce à analyser est pure et qu'aucune étape de sélection d'ions n'est nécessaire avant l'irradiation par des électrons) en font une technique de choix pour l'analyse des toxines. Néanmoins, nos expériences montrent également les limitations de cette approche. La première est liée à la quantité d'échantillon disponible (et à la complexité du mélange initial) puisqu'il faut absolument partir d'un signal intense pour l'ion parent, pour obtenir un spectre ECD informatif. Ce problème est moins

crucial sur les instruments FT-ICR hybrides puisque l'ion parent peut être concentré avant son entrée dans la cellule.

Une autre limitation est liée à la structuration que les protéines peuvent adopter en phase gazeuse. Si la protéine à fragmenter présente une structuration trop importante, l'ECD devient inefficace puisque les fragments formés restent liés ensemble par des interactions faibles. Idéalement une activation douce par IRMPD est souhaitable avant l'étape d'ECD.

D'un point de vue biologique, nous avons mis en évidence de nouvelles familles de toxines, aux propriétés originales pour certaines, pour l'ensemble des venins étudiés (α -conotoxines, conotoxines de la superfamille T, toxine spécifique des récepteurs α_{1A} -adrénergiques, sarafotoxines d'une taille originale).

En conclusion, nous avons, au cours de ce travail utilisé les différentes facettes de la spectrométrie de masse FT-ICR pour mettre au point des stratégies d'analyse efficaces et rapides de nouvelles toxines peptidiques. Les résultats que nous avons obtenus montrent tout l'intérêt d'utiliser cette technologie performante dans le domaine de la toxinologie, ce qui n'avait pour ainsi dire jamais été fait. Ils en montrent également les limitations et ouvrent des perspectives pour l'avenir. L'approche « top-down » sur protéines entières doit être optimisée et étendue à des systèmes de taille plus importante. L'installation d'un laser IR permettant d'activer les protéines avant l'ECD est en projet au laboratoire DCMR. Il est également envisagé d'appliquer l'approche « top-down » pour caractériser les zones d'interactions entre l'AdTx1 et son récepteur. L'idée est de marquer la zone d'interaction par des atomes de deutérium (par échanges H/D en solution) et d'analyser l'AdTX1 entière en phase gazeuse. Le déplacement vers des m/z plus élevés des différents ions fragments c et z après cet échange permettra d'identifier les régions portant des deutérium qui sont celles marquant la surface d'interaction.

Enfin, il convient de rappeler que ces résultats sont le fruit du dialogue qui s'est instauré entre notre équipe de spectrométristes de masse et nos collaborateurs biologistes (principalement au CEA Saclay).

Annexes

--	--

Annexe I : Partie instrumentale

I. Le spectromètre de masse FT-ICR APEX III de Bruker

Toutes les analyses ont été effectuées sur un spectromètre APEX III de Bruker (Brême, Allemagne) équipé d'un aimant supraconducteur de 7 Tesla et d'une cellule ICR de type Infinity Cell. Les sources d'ionisation disponibles sont l'électrospray, le nano-electrospray et le MALDI. Les aiguilles nano-electrospray (off-line) proviennent de Proxeon



(Odense, Denmark). Un spray stable est obtenu en appliquant un potentiel d'environ -700V entre l'aiguille et le capillaire de verre (métallisé à ses extrémités, Capillary sur la Figure 1) utilisé pour la désolvatation et le transfert des ions. Le débit estimé est de 20-50 nL.min⁻¹. Les ions formés sont stockés dans un hexapole pendant une seconde puis pulsés vers la cellule ICR par une série de lentilles électrostatiques. Les ions sont piégés par des potentiels de piégeage PV1 et PV2 respectivement de 0,9 et 0,95 V. Les spectres de masses sont généralement enregistrés dans une gamme de masse de m/z [500 - 2000] avec 512k points par spectre.

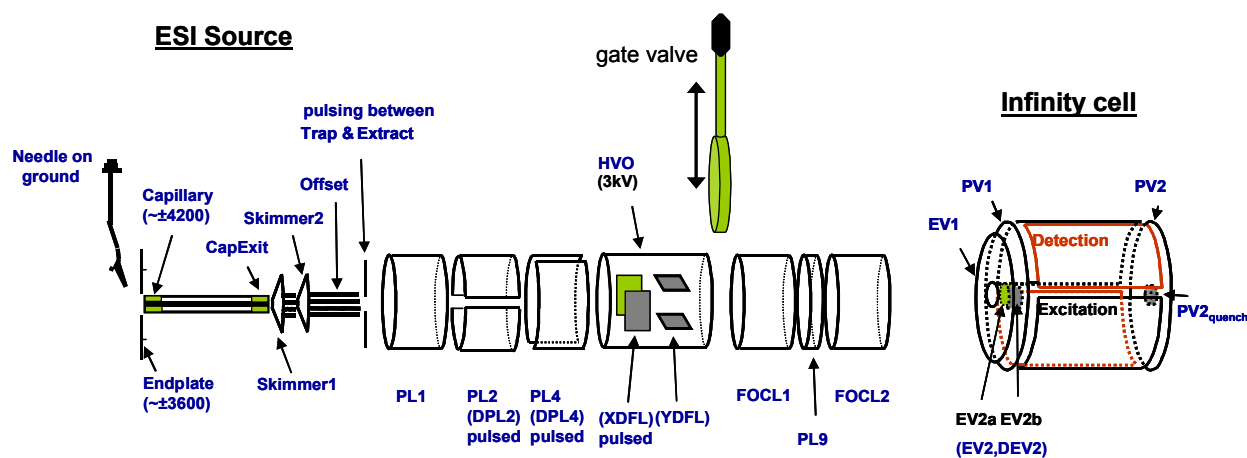


Figure 1: Schéma du FT-ICR APEX III Bruker, de la source électrospray à la cellule ICR Infinity Cell

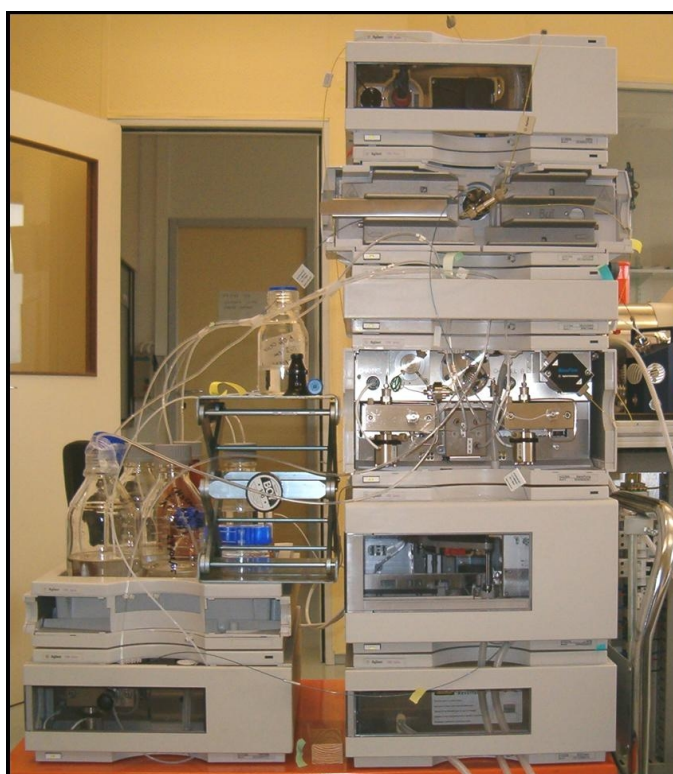
Le logiciel d'acquisition utilisé est Xmass de Bruker. Pour le retraitement des données plusieurs autres logiciels de Bruker ont été employés comme Biotools (séquençage *de novo*), Sequence Editor (éditeur de séquences peptidiques et protéiques) et DataAnalysis (traitement des chromatogrammes).

Pour les expériences de fragmentations par SORI-CAD, les ions parents sont isolés par éjection résonante sélective. Les fragmentations sont effectuées en utilisant un temps d'excitation de 400 μ s (V_{pp} =37dB) à une fréquence décalée de 500 Hz par rapport à la fréquence cyclotronique du parent. Le gaz de collision utilisé est du xénon. Pour les expériences d'ECD, les ions piégés sont généralement irradiés par un faisceau d'électrons de 1 eV pendant 250 ms.

La calibration (externe) du spectromètre pour les expérience d'infusion en nanoESI est effectuée quotidiennement par de l'iodure de sodium (NaI) à 2 mg.mL⁻¹ dans H₂O/iPrOH (50/50). Les clusters obtenus permettent d'étalonner la gamme de m/z 500 à 2000.

II. La chaîne de nanochromatographie liquide

La chaîne utilisée pour le couplage avec l'ICR est une chaîne Agilent Technologies 1100 series. La pré-colonne permettant le dessalage et la pré-concentration des échantillons est une Zorbax 300SB-C18 de 5 mm de longueur et de 0,3 mm de diamètre interne. Les particules de silice greffée octadecyle qui composent la phase stationnaire ont un diamètre moyen de 5 μ m. La colonne analytique est également une Zorbax 300SB-C18 de 15 cm de longueur, de diamètre interne 0,075 mm et dont le diamètre des particules de silice greffée est de 3,5 μ m.



Le diamètre interne de la pré-colonne est très élevé par rapport à celui de la colonne analytique afin d'obtenir une meilleure qualité de piégeage des composés. Le gradient

chromatographique utilisé est un gradient binaire constitué d'un solvant A (H₂O/AF, 99/1) et d'un solvant B (ACN/H₂O/AF 89,5/9,5/1).

Dans un premier temps, l'échantillon est acheminé de l'injecteur automatique à la précolonne à une vitesse de 10 $\mu\text{L}\cdot\text{min}^{-1}$ par une pompe isocratique (H₂O/AF 1%). L'échantillon piégé sur la pré-colonne est lavé pendant 5 min avec le solvant de cette pompe afin d'éliminer un maximum de sels. Ensuite, via un système de boucle à 6 voies, la pré-colonne est mise au contact de la pompe nanoLC qui va éluer les composés par un gradient de solvant B à un débit fixé à 300 $\text{nL}\cdot\text{min}^{-1}$. La composition initiale de sa phase mobile est de 5% de B pendant 5min (chargement de l'échantillon sur la pré-colonne) mais le gradient est variable en fonction de la complexité des échantillons à analyser. En fin de gradient, le système est toujours laissé plusieurs minutes à 100% de B afin de nettoyer son ensemble et d'éluer les composés très hydrophobes comme les polymères des plastiques. La colonne analytique est ensuite équilibrée avec la composition de solvant initiale.

III. Interface nanoLC / spectromètre FT-ICR

L'interface entre la nanoLC et le spectromètre est assurée par des aiguilles PicoTip™ de type SilicaTip™ de New objective (Woburn, MA, USA) possédant un diamètre externe de 360 μm , un diamètre interne de 75 μm et un orifice de 10 μm permettant l'émergence du spray. Les paramètres de formation du nanospray utilisés sont une tension de -1300 V appliquée sur la contre-électrode et une interface de désolvatation chauffée à 150°C.



Annexe II : Partie expérimentale

I. Analyse du venin d'*Atractaspis irregularis* par nanoESI-FT-ICR

La première étape consiste à prélever le venin (« traite » du serpent). Dans le cas d'*Atractaspis irregularis*, le serpent a été immobilisé dans un tuyau en plastique puis excité afin de le faire mordre dans un récipient recouvert d'un film polymérique (photo ci-contre). Le venin est récupéré puis lyophilisé. Après cette étape, le serpent est plongé dans l'azote liquide afin de pouvoir procéder au prélèvement des glandes à venin dont seront extraits les ARN messagers.



A. Analyse du venin brut par nanoESI

L'échantillon de venin brut lyophilisé (correspondant à environ 15 μ L de venin liquide) est repris dans 100 μ L d'eau acidifiée avec 1% d'acide formique (AF) pour former une solution stock. Pour chaque analyse, 1 μ L de cette solution stock est dilué 10 fois dans un mélange eau/méthanol/acide formique ($H_2O/MeOH/AF$), 49,9/49,9/0,2 % (v/v/v). 2 μ L de cette solution diluée, introduits dans une aiguille de nano-electrospray, suffisent pour réaliser de simples spectres MS (cartographies massiques). Pour les expériences de fragmentations, les capillaires sont remplis avec 5 μ L, ce qui permet de réaliser des expériences plus longues en accumulant plus de spectres. Entre 1/200 et 1/500^{ème} de l'échantillon est donc utilisé à chaque analyse.

B. Réduction des ponts S-S

Les ponts disulfure des toxines présentes dans le venin sont réduites *in situ* par le tris(2-carboxyethyl)phosphine (TCEP)¹. Le TCEP est un agent réducteur non-volatile plus stable et efficace que le dithiothréitol². Il réduit les ponts disulfures rapidement dans l'eau à

pH acide (pH 4,5). Un prélèvement de solution mère du venin brut est incubé avec un excès de TCEP. Plusieurs concentrations de TCEP ont été testées afin d'obtenir une réduction totale des ponts disulfure de cet échantillon. Il apparaît que le meilleur protocole est le suivant : 1 μ L de la solution mère est dilué avec 3 μ L d'un mélange H₂O/ACN/AF 49,5/49,5/1 (v/v/v) et incubé avec un microlitre d'une solution de TCEP à 1 nmol. μ L⁻¹. Le milieu réactionnel est placé sous agitation pendant 30 minutes à 50°C. L'échantillon réduit est enfin dilué avec 15 μ L du solvant de nano-electrospray H₂O/ACN/AF 49,5/49,5/1 (v/v/v). Ce protocole permet d'analyser l'échantillon directement après la réduction, sans purification supplémentaire, évitant ainsi la perte de composés qui pourrait être occasionnée par les étapes de purifications.

C Clonage des ARN messagers

Les ARNm ont été extraits et purifiés à partir des glandes à venin de l'unique spécimen *d'Atractaspis irregularis*, grâce à un kit de purification QuickPrep mRNA (Amersham, Biosciences). La banque d'ADN complémentaire a été synthétisée par RT-PCR (Ready-to-Go, Amersham Biosciences) par des amorces spécifiques aux sarafotoxines résultant de l'étude effectuée sur *A. engaddensis*. Ces amorces correspondent aux séquences complémentaires des régions non codantes (en 3' et 5') qui entourent les séquences des SRTXs matures. Les fragments d'ARNm purifiés résultant de la PCR ont été clonés en utilisant un kit TA-cloning (Invitrogen) puis transformés dans des bactéries *Echerichia coli* VαF'. Enfin, la séquence nucléotidique des clones a été déterminée grâce à un kit de séquençage BigDye terminator cycle et un séquenceur ABI 310 (Applied Biosystems, FosterCity, CA).

II. Analyse des venins de cônes marins

A. Analyse du venin de *Conus virgo*

Quinze spécimens de *Conus virgo* ont été collectés dans les îles des Philippines. Les glandes à venin ont été récupérées, débitées en petits morceaux, extraites plusieurs fois avec de l'eau (5% AF), puis centrifugées. Le venin brut, correspondant au surnageant a été récupéré puis lyophilisé.

A.1. Analyse du venin en infusion directe par nanoLC-FT-ICR

Une faible quantité de cet extrait est diluée dans 5 mL d'acétate d'ammonium 10 mM (pH 5), constituant ainsi notre solution stock. Pour chaque analyse, 100 µL de cette solution sont évaporés et le résidu obtenu est repris par 10 µL d'eau acide (1% AF). Le premier essai d'analyse en nanoESI sur le venin brut par simple dilution de la solution stock dans du méthanol ne permet pas l'obtention d'un spray, caractérisant la présence probable de sels dans l'échantillon. Une purification sur micro colonne de type ZipTip® a donc été introduite au niveau de la préparation d'échantillon permettant l'obtention d'un spray stable instantanément. Le dessalage des 10 µL de prise d'essai a été effectué en suivant le protocole suivant :

- 1- Lavage et conditionnement du ZipTip® avec un large volume de MeOH (ou ACN).
- 2- Rinçage du méthanol excédentaire par de l'eau acide (1% AF)
- 3- Adsorption de l'échantillon par une série de 15-20 aspirations-refoulements
- 4- Rinçage de l'échantillon par une dizaine d'aspirations-refoulements de 10 µL d'eau acide (1% AF)
- 5- Elution des composés piégés par un mélange d'eau, de méthanol et d'acide formique (H₂O/MeOH/AF) dans des proportions volumiques de 49,9/49,9/0,2.

Comme pour *Atractaspis irregularis*, 2 µL de cette solution, suffisent pour réaliser les spectres MS (cartographies massiques). Pour les expériences de fragmentations, les capillaires sont remplis avec 5 µL.

A.2. Analyse du venin réduit en infusion directe par nanoESI-FT-ICR

Cette réduction a été effectuée par le tris(carboxyethyl)phosphine (TCEP) directement sur le venin brut. 10 µL de solution mère concentrée ont été incubés avec 2 µL d'une solution de TCEP à 10 nmol.µL⁻¹ pendant 30 min à 50°C (pH=4). L'avancement de la réduction a été suivi par MALDI en déposant directement 0,5 µL du milieu réactionnel sur la cible MALDI puis en ajoutant 0,5 µL de matrice (DHB à 50 mg.mL⁻¹).

A.3. Conditions chromatographiques de l'analyse par nanoLC-FT-ICR

2 μ L de la solution mère concentrée sont injectés dans le système chromatographique. Le gradient analytique est fixé initialement à 5% de solvant B pendant 5 min. Une pente de 1% de solvant B par minute est imposée jusqu'à l'obtention d'une composition de phase mobile de 50% en solvant B. Cette pente de gradient constitue un bon compromis entre une bonne séparation et un temps d'analyse raisonnable. Il est généralement rapporté que peu de composés (peptides et protéines) sont élués avec des compositions de phases mobiles supérieures à 50% en phase organique. Aussi, dès qu'une composition de 50% en solvant organique est atteinte, la pente de gradient est augmentée à 4% par minute de manière à accélérer l'élution des derniers peptides ainsi que laver la colonne d'éventuels polymères introduits avec l'échantillon. Après un passage de 10 min à 100% de solvant B, cette dernière est reconditionnée à 5% en solvant B pendant un minimum de 30 min.

B. Analyse du venin de *Conus ermineus*

B.1. Analyse en infusion par nanoESI-FT-ICR

B.1.a. Venin natif

La dilution du venin brut de *Conus ermineus* dans le solvant d'électrospray ne permet pas, dans ce cas, l'observation d'un signal. Le probable présence de sels contraint à introduire une étape de purification sur micro colonne de type ZipTip®. Environ 5% de l'échantillon a été purifié de cette façon. Les composés piégés ont été élués par 10 μ L d'un mélange de H₂O/ACN/AF 49,95/49,95/0,1 % puis dilué 10 fois dans ce même solvant.

B.1.b. Analyse des fractions Er1A à Er1J

2 μ L de chacune des fractions K et D ont été purifiés sur microcolonne ZipTip® C₁₈ et élués par 10 μ L d'un mélange H₂O/MeOH/AF (50/50/50 ; v/v/v) pour pouvoir être analysés directement par nanoESI-FT-ICR.

B.2. Analyse en couplage nanoLC-FT-ICR

B.2.a. Venin natif

1% de l'échantillon de venin brut disponible (environ 3 μ g) est dilué 10 fois dans H₂O/AF (1%) puis 4 μ L de cette solution sont injectés dans le système chromatographique. La composition initiale de la phase mobile est de 5% de B pendant 5min (chargement de l'échantillon sur la précolonne). Au bout de 5 min, une pente de 0,5% B.min⁻¹ est appliquée jusqu'à ce que la composition en B atteigne les 50%. Cette pente très faible a pour but de séparer au maximum les composés du venin, le compromis est que le temps de l'analyse est long (2 heures environ). A partir de 50% de B, la pente de gradient est augmentée fortement à 5% B.min⁻¹ jusqu'à l'obtention de 100% de B. Cette composition est maintenue pendant 20 minutes afin de nettoyer le système de tous les composés fortement hydrophobes comme les polymères résiduels. La colonne est enfin équilibrée à 5% B pendant 30 min avant de pouvoir analyser un nouvel échantillon.

B.2.b. Venin réduit

4% de l'échantillon sont dilués par 36 μ L dans H₂O/AF (1%). L'échantillon est mis à réagir 30 minutes à 50 °C en présence de quelques cristaux de tris(carboxyethyl)phosphine (TCEP). 4 μ L de l'échantillon sont injectés dans le système chromatographique et sont analysés par nanoLC avec les mêmes conditions que celles du venin brut.

B.2.c. Synthèse de la toxine Er₁D₁₇₇₂

La synthèse de la toxine a été effectuée en phase solide en utilisant un synthétiseur ABI 430A (Applied Biosystems Division). La purification du peptide linéaire a été réalisée par RP-HPLC. La toxine est ensuite repliée par la formation de ses ponts disulfures dans les conditions oxydantes suivantes : pour 5 μ M de toxine, on ajoute 1 mM de glutathion réduit, 1 mM de glutathion oxydé, et 25% de glycérol dans un tampon Tris-HCl (pH 8). Le milieu réactionnel est laissé 48h à 4°C. Le peptide majoritaire est purifié par RP-HPLC et sa pureté est contrôlée par spectrométrie de masse.

III. Analyse du venin de *Dendroaspis angusticeps*

A. Analyse de la toxine naturelle DA₁EOA

A.1. Analyse directe par nanoESI-FT-ICR

Environ 600 ng de fraction purifiée DA₁EOA étaient disponibles pour non seulement effectuer une mesure de masse précise mais aussi dénombrer le nombre de ponts disulfures. Ceci correspond à environ 80 pmol de matière disponible (pour une masse de 7000 Da). Sur ces bases, la moitié de l'échantillon a été prélevée pour la première analyse. Ce prélèvement a été purifié sur ZipTip® C₁₈ et la toxine éluée de la microcolonne par 10µL d'un mélange H₂O/ACN/AF (49,95/49,95/0,1, v/v/v).

A.2. Analyse de la toxine réduite par nanoESI-FT-ICR

Pour obtenir une réduction des ponts totale et rapide, 10µL de solution mère ont été mis en présence d'un cristal de TCEP et le milieu réactionnel a été porté à 50°C pendant 30 min. La solution est ensuite purifiée sur Zip-Tip® C₁₈ et la toxine éluée par 10µL d'un mélange ACN/H₂O/AF (49,95/49,95/0,1 ; v/v/v).

A.3. Digestion enzymatique de DA₁EOA et analyse par infusion en nanoESI-FT-ICR

Pour cette expérience, environ 1,4 nmol (soit 10µg) de toxine DA₁EOA ont été digérés par 27pmol de trypsine dans un tampon bicarbonate d'ammonium (pH=7,9). Le rapport trypsine/toxine, classique, est de l'ordre de 1/50. La trypsine, qui permet le clivage en position C-terminale des deux acides aminés arginine et lysine, est choisie car la dégradation d'Edman a révélé la présence de nombreux résidus de ce type. Après digestion, la toxine est réduite par un excès de TCEP. Le fait de n'avoir ni réduit, ni alkylé la toxine avant la digestion permet d'obtenir des conditions de digestion ménagée puisque certains sites de clivages ne sont pas accessibles à l'enzyme. Ces conditions ont pour but d'obtenir une carte peptidique simple, composée de peu de peptides. Environ 50 pmol du mélange de peptides réduits sont purifiés sur ZipTip® C₁₈. Les peptides sont élués directement par 10µL de solvant d'électrospray (H₂O/ACN/AF, 49,95/ 49,95/ 0,1%) et analysés par infusion directe en nanoESI-FT-ICR.

B. Analyse de la toxine de synthèse AdTx1

B.1. Synthèse peptidique et repliement des ponts

La synthèse de la toxine a été effectuée sur phase solide en utilisant le synthétiseur ABI 430A (Applied Biosystems Division). La purification du peptide linéaire est effectuée par RP-HPLC. Les ponts disulfures de la toxine synthétique sont formés en la plaçant dans les conditions oxydantes suivantes : pour 2 μ M de toxine, on ajoute 1 mM de glutathion réduit, 1 mM de glutathion oxydé, et 20% de glycérol dans un tampon Tris-HCl (pH 8). Le milieu réactionnel est laissé 48h à 4°C. Dans ces conditions, il est couramment observé que le branchement des ponts disulfures formés est majoritairement identique à celui des toxines natives. Néanmoins, étant donné que différentes formes peuvent se former minoritairement (toxines oxydées partiellement, mauvais appariement des cystéines), une nouvelle purification par RP-HPLC est nécessaire. Le pic chromatographique très majoritaire caractérisant la toxine est collecté.

Une vérification du bon arrangement des ponts disulfures peut être effectuée par chromatographie à polarité de phases inverse. En effet, deux toxines de séquences identiques mais qui ne possèdent pas la même organisation des ponts disulfures auront des conformations tridimensionnelles différentes, donc des échanges différents avec la phase stationnaire ce qui se traduira par des temps de rétention différents. En comparant le temps de rétention de la toxine native à celui de la toxine synthétique, il est possible de vérifier si les deux toxines possèdent le même comportement hydrophobe (*Figure 2*).

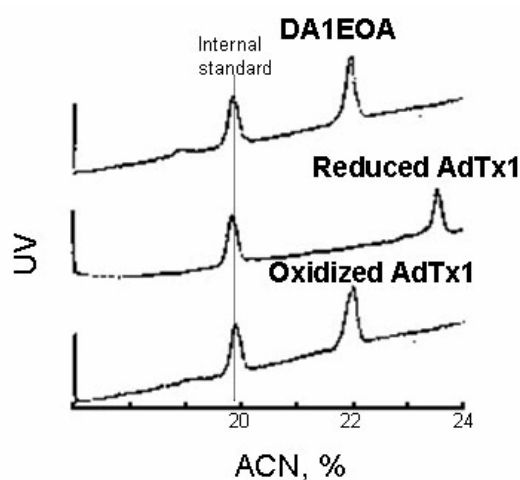


Figure 2: Comparaison des comportements hydrophobes de la toxine native DA1EOA et de la toxine synthétique AdTx1

Les résultats montrent que les deux toxines oxydées ont exactement le même temps de rétention, donc le même comportement hydrophobe. Ceci suggère qu'elles possèdent le même agencement des ponts disulfures.

B.2. Réduction des ponts disulfure du mutant

La réduction des ponts disulfure du ligand fluorescent a été effectuée en suivant le protocole suivant : 20µL de la fraction de purification de la toxine mutante sont réduits par l'ajout d'un cristal de TCEP. Le milieu réactionnel est placé 30 min à 50°C. L'AdTx1 mutée réduite est purifiée sur ZipTip® C₁₈ et l'élution effectuée avec 10µL d'un mélange H₂OMeO H/AF (49,9/49,9/0,2) permettant l'analyse directe par nanoESI-FT-ICR.

B.3. Marquage par le fluorophore Cy3B

Ce réactif est spécifique des amines primaires. Sur l'AdTx1(R*), cinq sites sont accessibles après la protection de la lysine 34 : 4 lysines et le groupement N-terminal. Les conditions utilisées pour modifier la toxine sont les suivantes 50 nmol de AdTx1(R*) oxydée sont mis en présence d'un excès de Cy3B (250 nmol) dans un tampon phosphate (0,3 mol.L⁻¹) à pH 7,1 pendant deux heures.

Annexe III : Méthodes d'analyse biochimiques

I- La dégradation d'Edman

A. Principe et processus chimique de la technique

Ce processus chimique permet, par étapes successives, de déterminer la séquence d'un peptide (ou d'une protéine) à partir de son extrémité N-terminale (*Figure 3*).

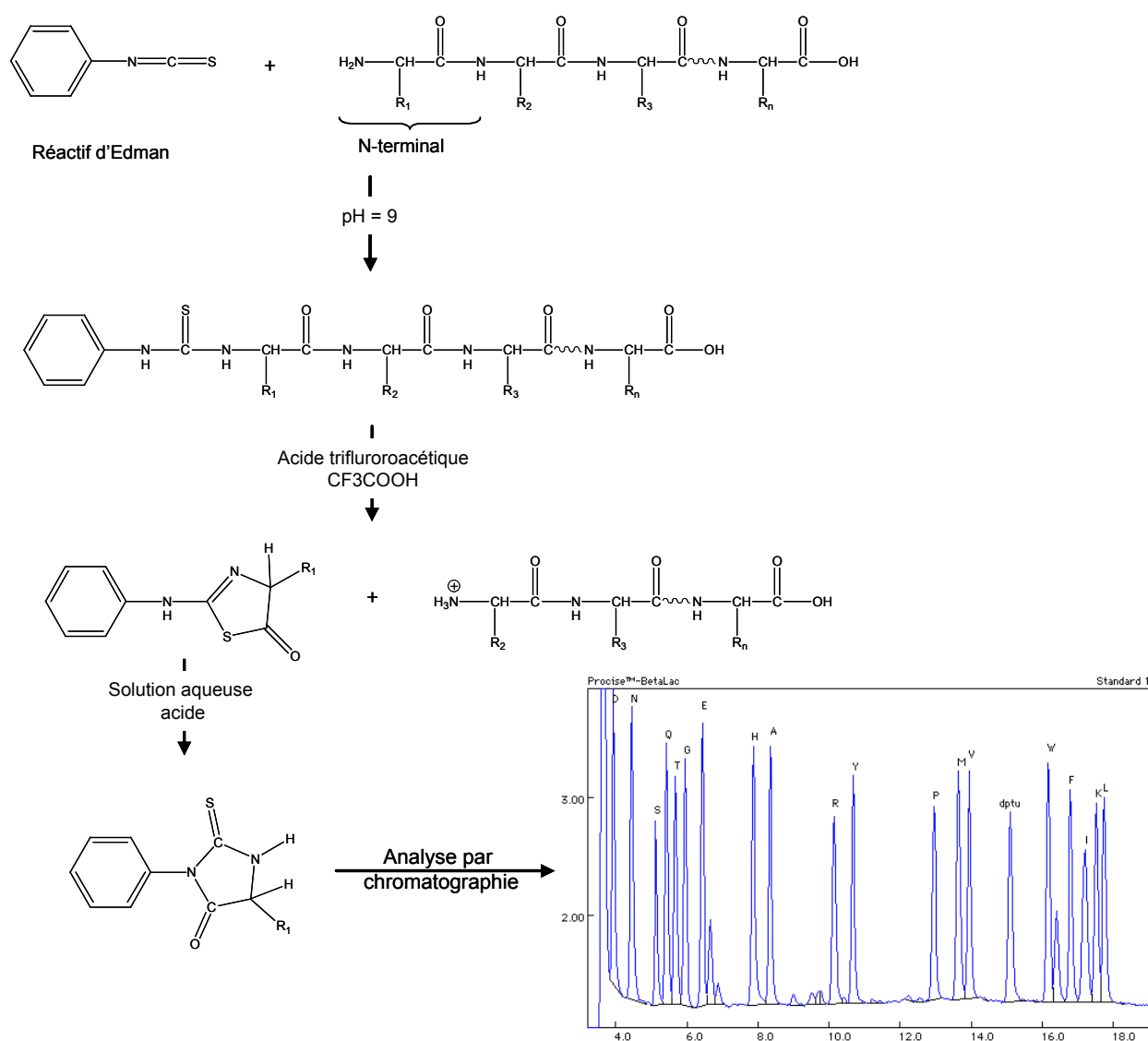


Figure 3: Principe de la dégradation d'Edman des peptides et protéines

A pH basique (pH=9), l'amine terminale du peptide réagit avec le réactif d'Edman (isothiocyanate de méthyle) pour former le dérivé phénylthiocarbamyle (PTC) du peptide. Le grand intérêt de ce réactif est qu'il est très spécifique de l'amine terminale, évitant ainsi de nombreuses réactions secondaires qui pourraient être envisagées avec les acides aminés basiques.

Ensuite, le dérivé PTC est clivé par l'acide trifluoroacétique afin de libérer d'un côté, un dérivé cyclique anilinothiazolinone (ATL) porteur de l'acide aminé N-terminal, et de l'autre côté, le peptide réduit à (n-1) acides aminés. La séparation entre ces deux composés se fait généralement par extraction liquide-liquide en utilisant comme solvant organique le chlorobutane ou l'acétate d'éthyle. Le dérivé ATL, plus hydrophobe est transféré dans la phase organique alors que le peptide, plus hydrophile reste en phase aqueuse. Pour éviter cette étape de purification délicate, une autre approche basée sur la fixation du peptide sur une phase solide *via* son extrémité C-terminale, permet de récupérer le dérivé ATL en phase liquide alors que le peptide (n-1), toujours greffé sur phase solide, peut être facilement remis en conditions pour le cycle suivant.

Le peptide à (n-1) acides aminés, est reconditionné pour subir un nouveau cycle de dégradation (pH 9 et présence d'un nouvel équivalent du réactif d'Edman). De son côté, le dérivé ATL est transformé en dérivé phenylthiohydantoïne (PTH), plus stable. Le PTH, porteur du groupement R_1 caractéristique du premier acide aminé, est ensuite analysé par chromatographie à polarité de phase inverse. Les conditions de gradient mises en place doivent permettre d'attribuer un acide aminé à un temps de rétention. Ceci est possible en jouant sur l'influence du groupement R_1 qui, en fonction de son hydrophobicité, induira un temps de rétention différent. Chaque acide aminé dégradé est analysé et identifié de la même manière. Un nombre n de cycles de dégradation permet donc (en théorie) de séquencer une protéine de n acides aminés et même de discriminer les acides aminés isomères leucine et isoleucine. Cette technique comporte malgré tout certaines limitations s'opposant à la théorie.

B. Limitations de la technique

Les premières limitations reposent sur la nature même de l'échantillon à séquencer. En effet, afin de limiter au maximum les réactions secondaires susceptibles de fausser les résultats, l'échantillon devra être très pur, ce qui est parfois délicat à obtenir. Un mélange de protéines sera impossible à séquencer correctement puisque plusieurs dérivés

phenylthiohydantoïne seront analysés en même temps en HPLC. La taille de la protéine va jouer un rôle non négligeable au niveau de la quantité d'échantillon à fournir pour optimiser les résultats. Même si quelques centaines de picomoles peuvent suffire pour séquencer une protéine de 50 kDa, les réactions secondaires dues aux conditions expérimentales atténuent le signal obtenu au fur et à mesure des cycles.

D'autres limitations de la technique sont liées à la séquence du peptide ou de la protéine à séquencer. Comme nous l'avons vu précédemment, la dégradation d'Edman est basée sur la réactivité de l'amine *N*-terminale. Aussi, elle ne pourra pas fonctionner sur les peptides qui présentent une modification en position *N*-terminale, comme une pyroglutamylation, une acétylation ou une méthylation, modifications courantes avec les toxines. De plus il est établi que les sérines et les thréonines forment des dérivés PTH peu stables ce qui conduit à des difficultés d'identification. D'autres acides aminés posent problème, les prolines et les glycines résistent partiellement à l'étape de clivage à l'acide trifluoroacétique. Pour exemple, le rendement de clivage des prolines pour cette étape est proche de 60%. Les 40% restant qui n'ont pas été clivés prendront part au cycle de dégradation suivant en introduisant de nombreux signaux parasites. Enfin, les cystéines impliquées lors de la présence de ponts disulfures, ne se retrouvent pas en phase liquide puisque le pont S-S la maintient au squelette peptidique. Il n'y a donc aucun dérivé analysable et le chromatogramme équivalent ne donne aucun signal, créant une nouvelle ambiguïté. Une alternative performante est de réduire et d'alkyler les cystéines préalablement à la dégradation d'Edman. Les cystéines, plus appariées par des ponts disulfures, peuvent être analysées par chromatographie en fin de cycle.

Du fait de ces différentes observations, il apparaît que la dégradation d'Edman ne permet pas de séquencer totalement la protéine dans un certain nombre de cas. Dans ce contexte, l'utilisation de la spectrométrie de masse peut s'avérer essentielle.

II. Le clonage moléculaire

A. De l'ADN à la synthèse des protéines chez les eucaryotes

L'ADN, acide désoxyribonucléique, est un polymère de 2'-désoxyribonucléotides liés entre eux par des liaisons phosphodiester. Chaque nucléotide est constitué par l'association d'une base (adénosine, guanine, thymine et cytosine), d'un sucre et d'un ou plusieurs groupe(s) phosphate(s). L'information génétique est portée par deux brins d'ADN assemblés de façon antiparallèle.

A.1. La transcription de l'ADN en ARN messagers chez les eucaryotes

La transcription de l'ADN est le processus au cours duquel l'information génétique contenue dans un gène va être transcrite en ARN. Cette transcription s'effectue dans le noyau grâce à des enzymes appelées ARN polymérases qui se lient sur une zone particulière de l'ADN nommée « promoteur » (*Figure 4*).

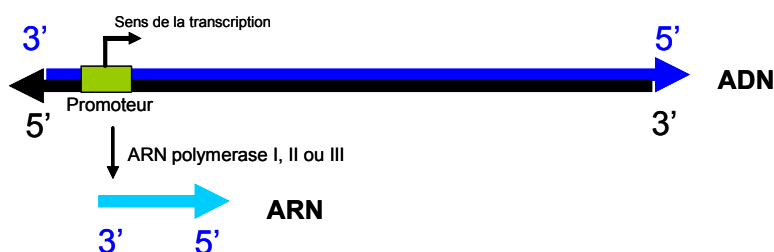


Figure 4: Transcription simplifiée de l'ADN en ARN

Les ARN messagers (ARNm) sont issus de la maturation des pre-ARNm obtenus à partir de l'ADN génomique par l'ARN polymérase de type II. Avant la fin de la transcription, les pre-ARNm sont coiffés sur leur extrémité 5', c'est-à-dire qu'ils vont subir une première modification. Cette modification consiste en l'addition enzymatique d'une guanosine qui va former un pont 5'-5' triphosphate avec le premier nucléotide transcrit (*Figure 5*). Elle va ensuite être méthylée sur l'azote de la position 7 de la guanine. Cette coiffe (ou *Capping*) va protéger les ARN d'éventuelles attaques enzymatiques tout en jouant un rôle important lors de la traduction. Cependant, elle ne constitue pas la seule modification que vont subir les pre-

ARNm. La seconde modification va avoir lieu après la transcription et consiste en l'addition d'une séquence de 100 à 200 résidus d'acide adénylique. Cette séquence nommée polyA est nécessaire pour que les ARNm soient reconnus et traduits efficacement par les ribosomes une fois dans le cytoplasme.

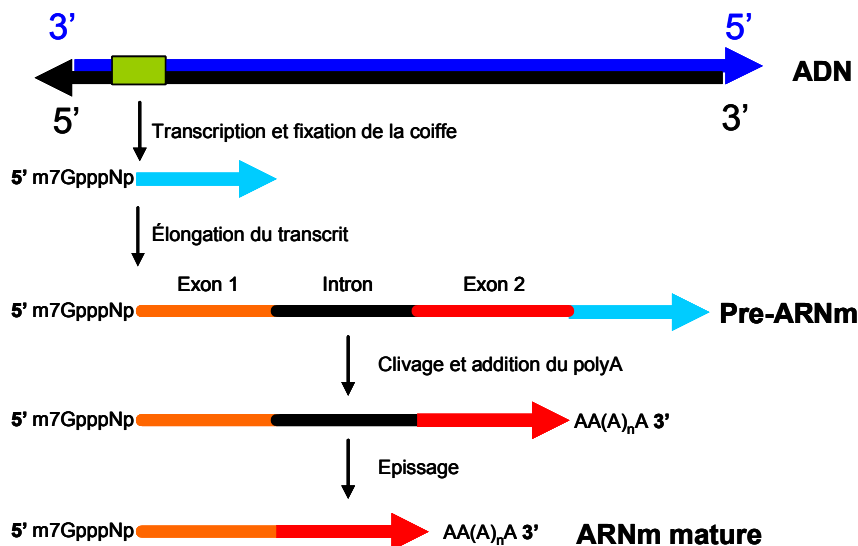


Figure 5: Transcription détaillée de l'ADN en ARN chez les eucaryotes

Les pre-ARNm transcrits possèdent de longues séquences non codantes, les introns, qui sont intercalés entre les séquences codantes, les exons. Ces introns doivent impérativement être éliminés pour que l'ARNm soit sous sa forme mature, c'est-à-dire pour qu'il soit prêt à être transporté vers le cytoplasme pour y être traduit. Cette élimination s'effectue dans le noyau au cours de l'épissage³, effectué par une machinerie très complexe, l'épissosome. Cette étape très complexe ne sera pas détaillée ici. Les ARNm résultant de cette transformation sont alors dit matures et prêts à être exportés du noyau vers le cytoplasme où ils seront traduits en séquences d'acides aminés.

A.2. La traduction des ARNm en protéines

Comme détaillé précédemment, l'extrémité 5' des ARNm est coiffée. Cette coiffe va permettre la reconnaissance de l'ARNm par une petite structure du ribosome (40S) et un complexe protéique. Ce complexe de préinitiation va alors scanner l'ARN en direction de son extrémité 3'. D'une manière générale, on peut considérer que la traduction va réellement commencer au premier codon AUG situé après la coiffe. Ce codon correspond à une méthionine qui est dite « méthionine d'initiation ». Ce signal va se traduire par la dissociation

du complexe protéique initial et par l'association avec une grosse sous-unité du ribosome nommée 60S. A partir de cet instant, chaque codon, ou triplet de nucléotides rencontré va donner lieu à l'émergence d'un signal de synthèse d'un acide aminé en suivant les règles du code génétique et via l'intervention des ARN de transfert. Cette étape de traduction va se prolonger jusqu'à la rencontre d'un codon « stop » dont les enchaînements de nucléotides sont UAA, UAG ou UGA.

Le clonage moléculaire permet de convertir et d'amplifier les ARNm purifiés de cellules spécifiques en séquences d'ADN complémentaires double-brin (ADNc). L'un des procédés utilisés est détaillé ci-dessous.

B. Principe du clonage moléculaire

Plusieurs protocoles de clonage moléculaire peuvent être utilisés. D'une manière générale, le clonage comporte les étapes suivantes : construction d'une banque d'ADN complémentaires (ADNc), amplification des ADNc d'intérêt, clonage dans des vecteurs plasmidiques, et séquençage de leur séquence nucléotidique.

B.1. Fabrication et amplification de la banque d'ADN complémentaire

La première phase consiste en la fabrication d'une banque d'ADN complémentaires (ADNc) à partir des ARNm extraits et purifiés des glandes à venin. Les ARNm sont purifiés par affinité par exemple par des billes magnétiques recouvertes d'amorces nucléiques d'oligo d(T) complémentaires des séquences poly (A) qui caractérisent les ARNm eucaryotes. Ces ARNm une fois purifiés sont ensuite convertis en copies d'ADN. Cette synthèse est catalysée par des transcriptases inverses qui sont des enzymes ADN polymérases ARN dépendantes. Elles sont capables d'utiliser un brin d'ARN pour catalyser la synthèse d'un brin d'ADN complémentaire de l'ARNm initial. Néanmoins, comme toutes les ADN polymérases, les transcriptases inverses ne peuvent pas initier seules la synthèse d'un brin d'ADNc (*Figure 6*). Elles ont besoin d'une séquence amorce. Lorsque que l'on part d'ARN messagers et dans la plupart des cas, l'amorce utilisée est une séquence de poly-thymines, qui va être complémentaire des poly-A greffés sur les ARNm au cours de la transcription. Tous les ARNm présents dans le milieu seront tout d'abord transformés en ADNc simple brin.

Classiquement, les molécules d'ADNc simple brin ainsi obtenues sont ensuite converties en ADNc double-brin *via* l'action combinée d'une enzyme Rnase de type H et d'une ADN polymérase. De manière alternative, il est possible de constituer une banque d'ADNc spécifique, c'est à dire correspondant à une sous-population particulière d'ARNm. Si on utilise des amorces oligonucléotidiques spécifiques de certains ARNm dont on connaît la structure et *via* la technologie d'amplification génique (PCR). On parle de RT-PCR (Reverse-Transcriptase-PCR). Le principe général de la PCR est décrit ci-dessous (Figure 7).

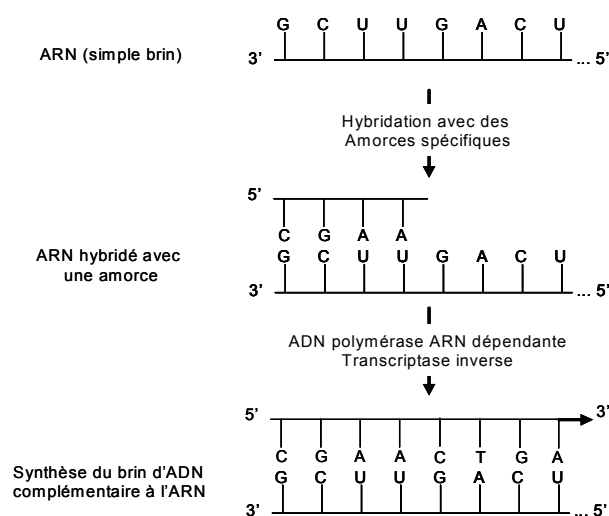


Figure 6: Transformation d'ARN en ADN complémentaire par l'intermédiaire de l'enzyme de transcription inverse (*reverse transcriptase*)

La première étape consiste à copier l'ARN en ADN *via* une transcriptase inverse mise en présence soit d'une amorce oligod(T), soit d'une amorce spécifique d'une séquence particulière de la population d'ARNm d'intérêt. Une seconde amorce complémentaire d'une région située à l'extrémité 5' de l'ADNc simple-brin néo-synthétisé, va permettre, par PCR, d'obtenir l'ADNc double-brin correspondant.

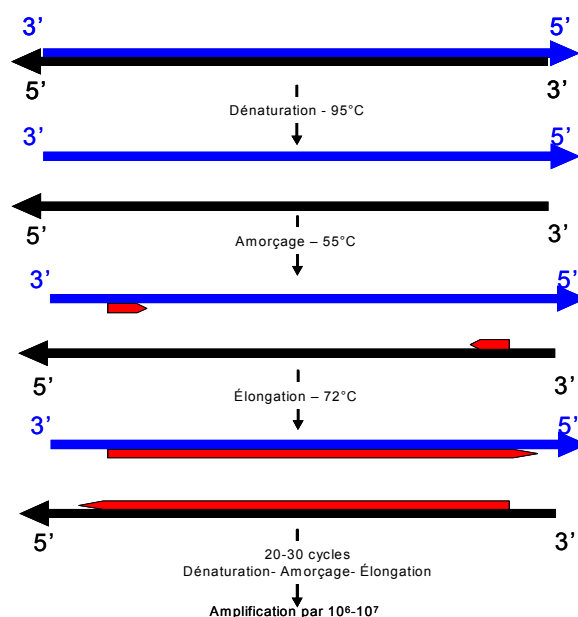


Figure 7: Principe de la réaction de polymérisation en chaîne

La réaction se décompose en trois étapes :

- 1- La dénaturation de l'ADN qui passe de sa forme double brin à sa forme simple brin. Elle s'effectue par une augmentation de la température du milieu réactionnel à 95°C pendant une minute.
- 2- L'amorçage : hybridation avec des amorces d'environ une vingtaine de nucléotides situées de part et d'autre du fragment à amplifier pendant 1 minute environ, à une température inférieure de 5°C de la température de fusion de l'amorce (T_m). Chaque amorce est complémentaire d'un des deux brins d'ADN initiaux.
- 3- L'élongation : une enzyme thermorésistante, la *taq* ADN polymérase va catalyser la synthèse d'un brin d'ADN dans le sens 5' \Rightarrow 3' à partir des extrémités 3' OH libres des amorces, et des quatre nucléotides triphosphate ajoutés dans le milieu réactionnel. Classiquement, cette synthèse s'effectue pendant 1 minute à 72°C.

L'amplification des ADNc se fait par une succession de cycles contenant ces trois étapes (classiquement on effectue 30 cycles successifs). Ainsi, les fragments synthétisés lors d'un premier cycle seront amplifiés au second et ainsi de suite. La théorie prédit une amplification exponentielle ce qui n'est pas toujours le cas expérimentalement. En général, on arrive à effectuer une vingtaine de cycles par heure, atteignant ainsi en théorie 2^{20} copies d'un ARN initial soit environ un signal amplifié 10^6 fois.

B.2. Clonage et séquençage de ADNc amplifié

Pour rendre le produit de PCR accessible au séquençage, il est cloné dans un vecteur plasmidique. La plupart des plasmides sont des molécules circulaires d'ADN double brin, beaucoup plus petites qu'un chromosome (3000 à 100000 paires de bases). La plupart des bactéries contiennent un ou plusieurs plasmides naturels. Ce sont des entités qui se répliquent en général de manière totalement indépendante du chromosome bactérien, tout en utilisant la machinerie de la cellule. Les ADN complémentaires que l'on souhaite cloner sont introduits dans des plasmides par ligation (*Figure 8*). Ce vecteur modifié est ensuite incorporé dans une bactérie qui, en se multipliant, va amplifier les plasmides et par voie de conséquence l'information contenue dans les ADNc.

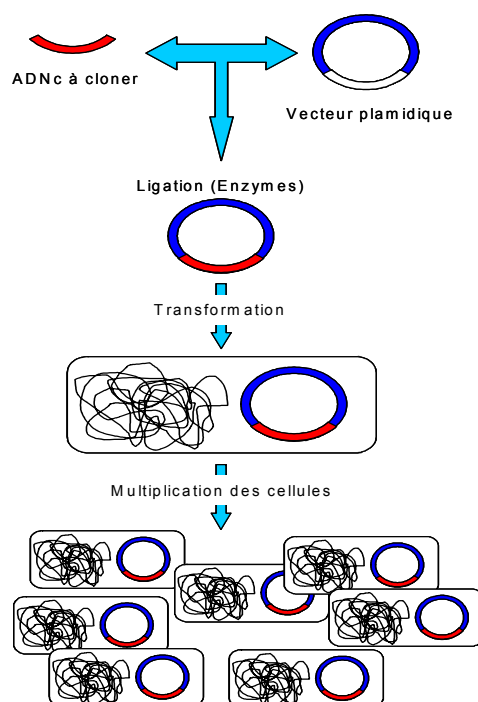


Figure 8: Clonage moléculaire dans un plasmide

Une fois que la multiplication est jugée satisfaisante (apparition de colonies), les bactéries sont récupérées et l'ADN plasmidique est purifié. Il peut ensuite être séquençé par la méthode de Sanger⁴. Les séquences nucléotidiques obtenues sont ensuite traduites en enchaînement d'acides aminés (par le code génétique) caractéristiques des peptides et protéines susceptibles d'être synthétisés.

Références bibliographiques

-
- ¹ Burn J.A., Butler J.C., Moran J., Whitesides G.M., Selective reduction of disulfides by Tris(carboxyethyl)phosphine, **1991**, *J. Org. Chem.*, 56, 2648-2650.
- ² Burmeister Getz E.L., Xiao M., Chakrabarty T., Cooke R., Selvin P.R., A comparison between the sulfhydryl reductants Tris(carboxyethyl)phosphine and dithiothreitol for use in protein chemistry, **1999**, *Anal. Chem.*, 273, 73-80.
- ³ Moore M.J., Query C.C., Sharp P.A., Splicing of precursors to mRNA by the spliceosome, **1993**, *The RNA World*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.p.p. 303-358.
- ⁴ Sanger F., Nicklen S., Coulson A.R., DNA sequencing with chain-terminating inhibitors, **1977**, *Proc. Natl. Acad. Sci.*, 74, 5463-5467.

Characterization of Toxins within Crude Venoms by Combined Use of Fourier Transform Mass Spectrometry and Cloning

Loïc Quinton,[†] Jean-Pierre Le Caer,[†] Gilles Phan,[‡] Caroline Ligny-Lemaire,[‡] Julie Bourdais-Jomaron,[§] Frédéric Ducancel,[‡] and Julia Chamot-Rooke^{*†}

Laboratoire des Mécanismes Réactionnels, UMR 7651 CNRS, Ecole Polytechnique, F-91128, Palaiseau, France, Département d'Ingénierie et d'Études des Protéines, Commissariat à l'Énergie Atomique, Centre d'Études de Saclay, Gif-sur-Yvette Cedex 91191, France, and Facultad de Ciencias Universidad Autonoma del Estado de Morelos, Av. Universidad 1001 Col. Chamilpa, Cuernavaca, Morelos 62210 Mexico

The standard analytical procedure for screening the proteomic profile of a venom often relies on an appropriate combination of sample extraction, electrophoresis, reversed-phase high-performance liquid chromatography, mass spectrometry, and Edman degradation. We present in this study a new approach for venom screening based on Fourier transform mass spectrometry (FTMS) analysis directly on the crude venom. The venom chosen is a unique sample from *Atractaspis irregularis*, a species never studied at the molecular level previously. This snake belongs to the Atractaspididae family that is known to produce highly toxic venoms containing endothelin-like peptides called sarafotoxins (SRTXs). Nanoelectrospray-FTMS spectrum of the crude venom allowed the identification of 60 distinct compounds with molecular masses from 600 to 14 000 Da, which would have been impossible without the resolution of this kind of instrument. De novo sequencing within the entire venom confirmed the sequences of two new families of sarafotoxins, whose precursors had been cloned, and allowed the characterization of a third one. One particularly interesting point was that the propolypeptides appeared processed not in one unique compound, but rather in different length molecules ranging from 15 for the shorter to 30 amino acids for the longer. Moreover, our results clearly establish that in the case of *A. irregularis* only one copy of mature sarafotoxin emerges from each precursor, which is a totally different organization in comparison of other precursors of SRTXs.

Venoms are highly complex mixtures that can contain several hundred compounds,¹ many of them having a potent biological activity. Even though most compounds are peptides (which can

be highly posttranslationally modified such as in the conus genus,^{2,3} proteins, enzymes (phospholipases),⁴ or small organic compounds such as polyamines⁵ are often also present in the same venom. The highly diverse biological activities observed for these different species allow venoms to be considered as a rich source of natural and active compounds. This diversity explains both the high toxicity of some of these venoms and the increasing number of venom screenings to discover new toxins.⁶

The standard analytical procedure for a screening of a complete protein profile relies on an appropriate combination of sample extraction, electrophoresis, reversed-phase high-performance liquid chromatography (RP-HPLC), mass spectrometry (MS),⁷ Edman degradation, and molecular cloning. The coupling between HPLC and MS can be either on-line or off-line.⁸ Two different kinds of information can be obtained from such experiments. The first is a mass fingerprint of the crude venom that can be further used for taxonomic identification.⁹ The second is the amino acid sequence (partial or total) of one toxin of interest using often de novo sequencing by mass spectrometry or Edman degradation. The mass spectrometric approach is of special interest when the peptide to be sequenced has its N-terminus blocked, which precludes any Edman degradation.¹⁰ Considering the complexity of the crude venom, a prior purification of the toxin of interest by different steps of RP-HPLC is generally needed before MS/MS experiments. One of the reasons is that electrospray ionization,

- (2) Olivera, B. M. *Mol. Biol. Cell* **1997**, *8*, 2101–2109.
- (3) Olivera, B. M.; Rivier, J.; Clark, C.; Ramilo, C. A.; Corpuz, G. P.; Abogadie, F. C.; Mena, E. E.; Woodward, S. R.; Hillyard, D. R.; Cruz, L. J. *Science* **1990**, *249*, 257–263.
- (4) Bao, Y.; Bu, P.; Jin, L.; Hongxia, W.; Yang, Q.; An, L. *Int. J. Biochem. Cell Biol.* **2005**, *37*, 558–565.
- (5) Tzouros, M.; Manov, N.; Bienz, S.; Bigler, L. *J. Am. Soc. Mass Spectrom.* **2004**, *11*, 1636–1643.
- (6) Diego-García, E.; Batista, C. V. F.; García-Gómez, B. I.; Lucas, S.; Candido, D. M.; Gómez-Lagunas, F.; Possani, L. D. *Toxicon* **2005**, *45*, 273–283.
- (7) Hansson, K.; Furie, B.; Furie, B. C.; Stenflo, J. *Biochem. Biophys. Res. Commun.* **2004**, *319*, 1081–1087.
- (8) Vianna Braga, M. C.; Konno, K.; Portaro, F. C. V.; de Freitas, J. C.; Yamane, T.; Olivera, B. M.; Pimenta, D. C. *Toxicon* **2005**, *45*, 113–122.
- (9) Escoubas, P.; Chamot-Rooke, J.; Stöcklin, R.; Whiteley B. J.; Corzo, G.; Genet, R.; Nakajima, T. *Rapid Commun. Mass Spectrom.* **1999**, *13*, 1861–1868.
- (10) Pereira Ribeiro, S.; Mendes, M. A.; Delazari dos Santos, L.; Monson de Souza, B.; Ribeiro Marques, M.; Filgueira de Azevedo, W., Jr.; Palma, M. S. *Peptides* **2004**, *25*, 2069–2078.

* Corresponding author: (phone) 33 1 69 33 34 07; (fax) 33 1 69 33 30 41. (e-mail) jcr@dcmr.polytechnique.fr.

[†] Ecole Polytechnique.

[‡] Centre d'Études de Saclay.

[§] Universidad Autonoma del Estado de Morelos.

(1) Kalume, D. E.; Stenflo, J.; Czerwiec, E.; Hambe, B.; Furie, B. C.; Furie, B.; Roepstorff, P. *J. Mass Spectrom.* **2000**, *35*, 145–156.

which is the most common ionization mode used, often yields an envelope of multicharged ions for each species that complicates the isolation of the peak of interest before its fragmentation. Another reason is that the dynamic range of most mass spectrometers is not sufficient to allow the less intense peaks in the mass spectrum to be characterized. A split of the crude venom into several less complex mixtures helps to overcome this problem. Obviously, one solution to avoid these fastidious chemical steps of purification would be to work directly within the crude venom. Such an approach on an entire and unpurified venom has been attempted recently with success by NMR spectroscopy, allowing the characterization of new sulfated nucleosides in the spider venom of *Tegenaria agrestis*.¹¹ A mass spectrometry based approach would necessitate the use of a mass spectrometer with a particularly high resolving power to discriminate all species of a complex mixture. Most mass spectrometers (quadrupoles, ion traps) do not fulfill this condition with the exception of Fourier transform mass spectrometers. Fourier transform mass spectrometry (FTMS)^{12–14} provides the highest performance in mass resolution and mass accuracy and for that reason is gaining more and more importance in analyzing complex mixtures obtained from protein digests,¹⁵ synthesized by combinatorial peptide chemistry,¹⁶ or simply derived from natural sources.¹⁷ The mass accuracy provided by such instrumentation can also be an important aid in de novo sequencing.^{18,19} Moreover, FTMS has some advantages over NMR with respect to mixture complexity, dynamic range, and size of compounds that can be analyzed. In their NMR study of crude venoms, Taggi et al. showed that the complexity of some venoms necessitates a prefractioning step by RP-HPLC before the NMR analysis.¹¹ They also showed that the experimental conditions chosen for HPLC are very important and that more than 50% of the material under investigation can be lost if standard chromatographic techniques are used, which emphasizes the interest of working on crude extracts. Indeed, NMR and FTMS can be considered as complementary techniques since NMR can provide valuable information on quantitation within the mixture or structural data that could be hardly obtained by mass spectrometry.

The venom chosen for this study is a unique sample from *Atractaspis irregularis*, a species never studied at the molecular level previously. This snake belongs to the Atractaspididae family, which is known to produce highly toxic venoms that contain endothelin-like peptides called sarafotoxins (SRTXs).²⁰ Recently, these vasoactive peptides were found to be 21 amino acids long

with 2 conserved disulfide bridges and to be produced from original precursors.²¹ Interestingly, a recent analysis of the venom of a specimen of *Atractaspis microlepidota microlepidota* revealed six new SRTX molecules longer at their C-terminus extremities by three amino acids than the previously described SRTXs.^{22,23} These studies relied on Edman degradation, mass spectrometry on HPLC purified toxins, and cloning of mRNA.

We show here that nanoelectrospray-FTMS is a very powerful tool to analyze complex mixtures such as crude venoms and that characterization of new toxins can be done within the mixture, i.e., without any chromatographic separation prior to the MS analysis. It is also demonstrated that the use of FTMS is necessary for this kind of approach. Nanoelectrospray was chosen because it is a very low sample consumption ionization technique, which was necessary because the total amount of available venom was only a few microliters. This study reveals the presence of another new family of sarafotoxin isoforms, some of them being longer by up to 9 amino acids than the 21-amino acid-long SRTXs. The structure of the precursors encoding two of these new SRTX isoforms will be presented, revealing a different organization of the mRNA as compared to the previous ones.

EXPERIMENTAL SECTION

Mass Spectrometry. All experiments were carried out with a 7-T APEX III FT-ICR mass spectrometer (Bruker Daltonik, Bremen, Germany) equipped with a 7-T actively shielded superconducting magnet and an infinity cell.²⁴ The ionization source used was the Apollo nanoelectrospray source. Nanoelectrospray needles were purchased from Proxeon (Odense, Denmark) and filled with 2 μ L of the sample diluted in 50:50:0.2 MeOH/water/formic acid (v/v/v) and subsequently opened by breaking the tapered end of the tip under a microscope. A stable spray was observed by applying a voltage of about -700 V between the needle (grounded) and the entrance of the glass capillary used for ion transfer. A source temperature of 50 °C was used for all nanoelectrospray experiments. The estimated flow rate was 20 – 50 nL/min. Ions were stored in the source region in a hexapole guide for 1 s and pulsed into the detection cell through a series of electrostatic lenses. Ions were finally trapped in the cell using SideKick and front and back trapping voltages of 0.9 and 0.95 V. Mass spectra were acquired from m/z 500 to 3000 with $512k$ data points, and monoisotopic peaks were automatically labeled using the software XMASS 6.1.4 (Bruker Daltonics). For sustained off-resonance irradiation collision-activated dissociation (SORI-CAD)²⁵ experiments, the parent ion was isolated by radio frequency (rf) ejection of all unwanted ions using both low-voltage single rf pulses (soft shots) at their resonance frequencies and a chirp excitation

(11) Taggi, A. E.; Meinwald, J.; Schroeder, F. C. *J. Am. Chem. Soc.* **2004**, *126*, 10364–10369.

(12) Marshall, A. G.; Hendrickson, C. L. *Int. J. Mass Spectrom.* **2002**, *215*, 59–75.

(13) Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S. *Mass Spectrom. Rev.* **1998**, *17*, 1–35.

(14) Marshall, A. G. *Physica B* **2004**, *346–347*, 503–508.

(15) Li, L.; Masselon, C. D.; Anderson, G. A.; Pasa-Tolic, L.; Lee, S.-W.; Shen, Y.; Zhao, R.; Lipton, M. S.; Conrads, T. P.; Tolic, N.; Smith, R. D. *Anal. Chem.* **2001**, *73*, 3312–3322.

(16) Schmid, D. G.; Grosche, P.; Bandel, H.; Jung, G. *Biotechnol. Bioeng. (Comb. Chem.)* **2001**, *71*, 149–161.

(17) Barrow, M. P.; Headley, J. V.; Peru, K. M.; Derrick, P. J. *J. Chromatogr., A* **2004**, *1058*, 51–59.

(18) Spengler, B. *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 704–715.

(19) Schlosser, A.; Lehmann, W. D. *Proteomics* **2002**, *2*, 524–533.

(20) Kloog, Y.; Ambar, I.; Sokolovsky, M.; Kochva, E.; Wollberg, Z.; Bdolah, A. *Science* **1988**, *242*, 268–270.

(21) Ducancel, F.; Matre, V.; Dupont, C.; Lajeunesse, E.; Wollberg, Z.; Bdolah, A.; Kochva, E.; Boulain, J. C.; Menez, A. *J. Biol. Chem.* **1993**, *268*, 3052–3055.

(22) Hayashi, M. A. F.; Ligny-Lemaire, C.; Wollberg, Z.; Wery, M.; Galat, A.; Ogawa, T.; Muller, B. H.; Lamthanh, H.; Doljanski, Y.; Bdolah, A. V.; Stöcklin, R.; Ducancel, F. *Peptides* **2004**, *25*, 1243–1251.

(23) Menin, L.; Favreau, P.; Doljanski, Y.; Ducancel, F.; Stöcklin, R. The proteome of *Atractaspis microlepidota microlepidota* snake venom reveals long-sarafotoxins, a family of endothelin-like peptides. *Proceedings of the 15th European Symposium on Animal, Plant and Microbial Toxins (Slovenia)*; 2004; p 68.

(24) Caravatti, P.; Allemann, M. *Org. Mass Spectrom.* **1991**, *26*, 514–518.

(25) Gauthier, J. W.; Trautman, T. R.; Jacobson, D. B. *Anal. Chim. Acta* **1991**, *246*, 211–225.

covering the region of interest. The fragmentation was performed using an excitation duration of 400 μ s ($V_{pp} = 4$ V) at a frequency 500 Hz higher than the cyclotron frequency of the ion of interest. Xenon was used as the collision gas and introduced by several pulses through a pulsed valve to obtain a pressure in the cell up to 10^{-7} mbar. The excitation/detection of all fragment ions was performed after a 4-s pumping delay. An external mass calibration using NaI (2 mg/mL in 2-propanol/water) was done daily.

cDNA Library Construction and Gene Sequencing. The poly(A)⁺ RNA fraction that corresponds to the messenger RNAs (mRNAs) was extracted and purified from the venom gland of a single *A. irregularis* specimen using a QuickPrep mRNA purification kit (Amersham, Biosciences). Complementary DNA (cDNA) were synthesized by reverse transcriptase polymerase chain reaction (RT-PCR) using the Ready-to-Go kit (Amersham Biosciences) and a set of SRTX specific primers complementary to 5' and 3' noncoding regions that surround the open reading frame (ORF) contained in the precursors encoding SRTXs from *Atractaspis engaddensis*: SRTX-5'nc 5'-GCCCCAYCTNCTNTGCTCV-GCSMG-3' (Y = C + T; S = G + C; M = A + C; V = G + A + C; N = A + C + G + T) and SRTX-3'nc 5'-GAGACTGGATGTCTTTCCTAGCTA-3'. The resulting purified-PCR DNA fragments were cloned and transformed in *Escherichia coli* VøF' strain using a TA-clonin kit (Invitrogen). The nucleotide sequence of the clones was determined using the BigDye terminator Cycle Sequence Kit and an ABI 310 automatic system (Applied Biosystems, Foster City, CA).

Sample Preparation. Lyophilized crude venom (from ~15 μ L of venom after snake milking) was dissolved in 100 μ L of acidic water (1% formic acid) to form a stock solution. One microliter of this stock solution was diluted in 9 μ L of 50:50:0.2 MeOH/water/formic acid (v/v/v), and 2 μ L of this final solution was used for each nanoelectrospray experiment. Therefore, for each analysis, 1/500 of the initial sample was consumed.

Reduction Procedure. Reduction of the crude venom was performed using tris(2-carboxyethyl)phosphine (TCEP).²⁶ TCEP is an odorless and nonvolatile reducing agent that is more stable and effective than dithiothreitol.²⁷ It reduces disulfides rapidly and completely in water at pH 4.5. One microliter of the crude venom stock solution was incubated with 1 μ L of TCEP (1 nmol/ μ L) at 50 °C for 30 min at pH 4. The course of the reduction was monitored from the shifts of ions to higher m/z in the mass spectrum (the reduction of each S-S bond yields a +2 Da shift for the molecular mass).

RESULTS AND DISCUSSION

Analysis of the Crude Venom. The direct analysis by nanoelectrospray-FTMS of the crude venom exhibits ~60 distinct molecular masses in the range 600–15 000 Da (Table 1). The corresponding mass spectrum is depicted in Figure 1.

It is worth noting that the resolution of FTMS is required to determine the charge state of all isotopic patterns obtained for the crude venom. Indeed, many species present in the spectrum depicted in Figure 1 appear in only one charge state, which means that their molecular mass is calculated on the basis of the charge

state obtained from the resolution of isotopic distribution and not from the distribution of multicharged ions. The resolution provided by an ion trap or a QTOF instrument would not be sufficient to resolve the isotopic patterns for highly charged ions.

The use of electrospray ionization enhances the problem of ion suppression. It is clear that the wide mass range of the different species present in a venom and the fact that these compounds may have various physicochemical properties emphasizes this problem. Nevertheless, it is worth noting that the use of nanoelectrospray instead of electrospray increases the ionization efficiency, due to a smaller initial droplet size, and thus reduces the ion suppression.^{28,29} Moreover, ion suppression would be a serious limitation if the aim of our study was to determine all classes of compounds (organic, biological, polar, nonpolar, basic, and acidic species) present in the venom for its complete characterization. But the objective of this study is not to identify all the species present in the venom but rather to get a mass fingerprint of the venom and look for a specific class of peptides (sarafotoxins) previously studied by mass spectrometry. For that purpose, the drawback of ion suppression is not so crucial.

Comparison with the crude venom of the previously studied species *A. microlepidoda microlepidoda* revealed a similar overall diversity of these venoms in that range of molecular masses.^{22,23}

One remarkable feature is that half of the masses are between 2000 and 3500 Da, which correspond to peptides between 20 and 30 amino acids, a characteristic length for sarafotoxins. Nevertheless, it is notable that none of the masses identified in this venom corresponds to a sarafotoxin isoform described previously in other *Atractapidae* venoms. This can be simply explained by a different composition of all the venoms of *Atractapidae*. Although the venoms share similarities in terms of mass fingerprinting, no common sarafotoxin was ever found between the different samples studied.

To further aid in the identification of the compounds that could correspond to new SRTX isoforms, we have undertaken a reduction of the crude venom with the objective of gaining insight into the disulfide bond content in all molecules present in the mixture (Table 1, Figure 2).

The shifts in molecular masses measured after reduction are reported in Table 1. Most of the masses (31 of 54) are displaced by 4 Da, indicative of the presence of two disulfide bonds, as expected for sarafotoxins. It is noteworthy that this shift in mass occurs primarily in molecules ranging from 1.9 to 3.5 kDa versus 2.5–2.9 kDa in the case of known SRTXs. It can also be seen that ions having a (3+) charge state are the major ones observed in the nanoelectrospray spectrum before TCEP reduction and remains the major one after the reduction. Nevertheless, some highly charged ions clearly disappear to give lower charge states. Actually, we also observed this general trend indicating a decrease of the charge state after the reduction procedure for other venoms and purified toxins. Since no cleanup was done after the reduction, a discrimination due to the purification step can be discarded. It is true that, for proteins, the reduction of S-S bonds yields an elongated structure and thus facilitates the protonation to sites that were either buried within the compact structure or neighboring another potential protonation site. Nevertheless, it

(26) Burn, J. A.; Butler, J. C.; Moran, J.; Whitesides, G. M. *J. Org. Chem.* **1991**, *56*, 2648–2650.

(27) Burmeister Getz, E. L.; Xiao, M.; Chakrabarty, T.; Cooke, R.; Selvin, P. R. *Anal. Chem.* **1999**, *273*, 73–80.

(28) Wilm, M.; Mann, M. *Anal. Chem.* **1996**, *68*, 1–9.

(29) Juraschek, R.; Dulcks, T.; Karas, M. *J. Am. Soc. Mass Spectrom.* **1999**, *10*, 300–308.

Table 1. Molecular Masses (before and after Disulfide Bond Reduction) Observed in the Venom of *A. irregularis* and Calculated after Deconvolution of the Nano-ESI FT-ICR Spectra

molecular mass (Da) before S-S reduction	intensity ^a	<i>M_r</i> after S-S reduction	no. of disulfide bonds ^b	molecular mass (Da) before S-S reduction	intensity ^a	<i>M_r</i> after S-S reduction	no. of disulfide bonds ^b
614.3035	+	614.3068	0	2994.2129	+	2998.2157	2
636.2991	+	636.3024	0	3002.1814	+	3006.2176	2
655.3374	+	655.3416	0	3004.1853	+	3008.2360	2
730.3340	+	730.3373	0	3010.1361	+	3014.1771	2
858.3933	+	858.3971	0	3018.2025	+	3020.1896	1
1419.6683	+	1421.6844	1	3085.2778	+	3089.3119	2
1833.5666	+	1837.6033	2	3293.3572	+	3297.3991	2
1929.8034	+	1933.8411	2	3353.4441	+	3357.4967	2
1935.6807	+	1939.7204	2	3515.4929	+	3519.5338	2
2027.7254	+	2031.7667	2	3537.5296	+	3541.5829	2
2050.7077	++	2054.7464	2	4606.8665	+	4606.8669	0
2076.6872	+	2080.7220	2	4633.8773	+	4633.8831	0
2175.7575	+	2179.7866	2	4703.9061	+	4703.9155	0
2280.8160	++	2284.8523	2	4730.9292	+	4730.9201	0
2288.8368	+	2292.8768	2	4774.9651	+	4774.9609	0
2296.8174	+	2300.8322	2	5392.3077	+	5394.2158	1
2474.9249	+	2478.9490	2	5421.1969	+	5421.2020	0
2680.0071	+	2684.0451	2	5490.5389	++	5498.5137	4
2688.0291	+	2692.0670	2	6976.9661	+	6985.0444	4
2702.0493	++	2706.0791	2	7613.3364	+	7620.4120	4
2926.2049	+	2928.2183	1	8343.9918	+	8352.0782	4
2940.2232	+	2942.2370	1	12917.4319	++	12921.5302	2
2950.1462	+++	2954.1864	2	12950.4199	+	12954.4713	2
2958.1769	++	2962.2160	2	12981.9043	+	12981.4783	0
2966.1492	++	2970.1921	2	12983.2015	+	12987.2137	2
2972.1857	+++	2976.2238	2	13274.3187	+	13280.4375	3
2982.1409	+	2984.1850	1	14215.2733	++	14231.5071	8
2988.1902	++	2992.1345	2				

^a Abundance of the species in the native spectrum is indicated by the number of “+”. +: intensity (*I*) of the ion compared to the most abundant ion, $0 < I < 5\%$. ++: intensity of the ion compared to the most abundant ion, $5 < I < 20\%$. +++: intensity of the ion compared to the most abundant ion, $20 < I < 100\%$. ^b Number of disulfide bonds for each species is the difference between molecular masses measured before and after S–S reduction divided by 2.

is difficult to compare proteins and peptides since it is not clear that for small molecules such as peptides the reduction of S–S bonds yields a really different structure where new sites of protonation become accessible. One possible explanation could be a higher proton affinity (PA) for the intact disulfide bond than for the free thiols. For instance, the PA of dimethyl disulfide ($815.3 \text{ kJ mol}^{-1}$) is more than 40 kJ mol^{-1} higher than the PA of methanethiol ($773.4 \text{ kJ mol}^{-1}$).³⁰ Another explanation could be a better solvation of the basic residues in the reduced species, which yielded a less efficient protonation in the gas phase.

To gain more insight into the sequences of the compounds, two additional types of experiment were undertaken: molecular cloning and de novo sequencing by mass spectrometry within the reduced crude venom. Indeed, the organization of sarafotoxins with two conserved disulfide bridges between cysteines +1/+15 and +3/+11 leads to a typical and unique Cys₁-X-Cys₃ ... Cys₁₁-X-X-Cys₁₅ signature among that family of bioactive peptides that precludes any de novo sequencing of the oxidized species.

Molecular Cloning of *A. irregularis* SRTX Precursors.

In Figure 3A is shown the structure of one of the two mRNA molecules that were cloned by homologous RT-PCR amplification using nucleic primers derived from the precursor of *A. engaddensis*

SRTXs.²⁰ To ensure the cloning of the mature sequence encoding mature SRTXs, the set of primers we designed are specific for two regions located upstream the signal peptide sequence and below the stop codon (Figure 3).

Owing to the position of the primers used for amplification, the two DNA fragments cloned contain a complete ORF of 409 base pairs encoding a precursor protein of 118 amino acid residues, where as expected, most of the 3' noncoding region was missing. Clearly, this corresponds to the first precursor of sarafotoxins ever described in the case of *A. irregularis* as illustrated by the presence of a SRTX sequence at its C-terminus extremity. The amino acid sequence deduced shows that the precursor protein begins with an initiating methionine followed by a hydrophobic region characteristic of a secretory signal peptide sequence whose cleavage site is predicted to be localized after the glycine residue at position 23 of the precursor polypeptide according to the predictive methods available.³¹ Indeed, it is now admitted that if signal peptides generally contain very few primary sequence constraints, the signal cleavage region is ideally six residues long, with neutral small side-chain amino acids at positions –1 (23rd amino acid residue of the ORF, Figure 3) and –3 (20th residue of the ORF, Figure 3) such as alanine, glycine, serine, and threonine. Following that predicted signal peptide and the beginning of the SRTX sequence, we note the presence of a

(30) Hunter, E. P.; Lias, S. G. Proton affinity evaluation. In *NIST Chemistry Webbook, NIST Standard Reference Database Number 69*; Lindstrom, P. J., Mallard, W., Eds.; National Institute of Standards and Technology: Gaithersburg, MD, 20899, 2003 (<http://webbook.nist.gov>).

(31) von Heijne, G. *Nucleic Acids Res.* **1986**, *14*, 4683–4690.

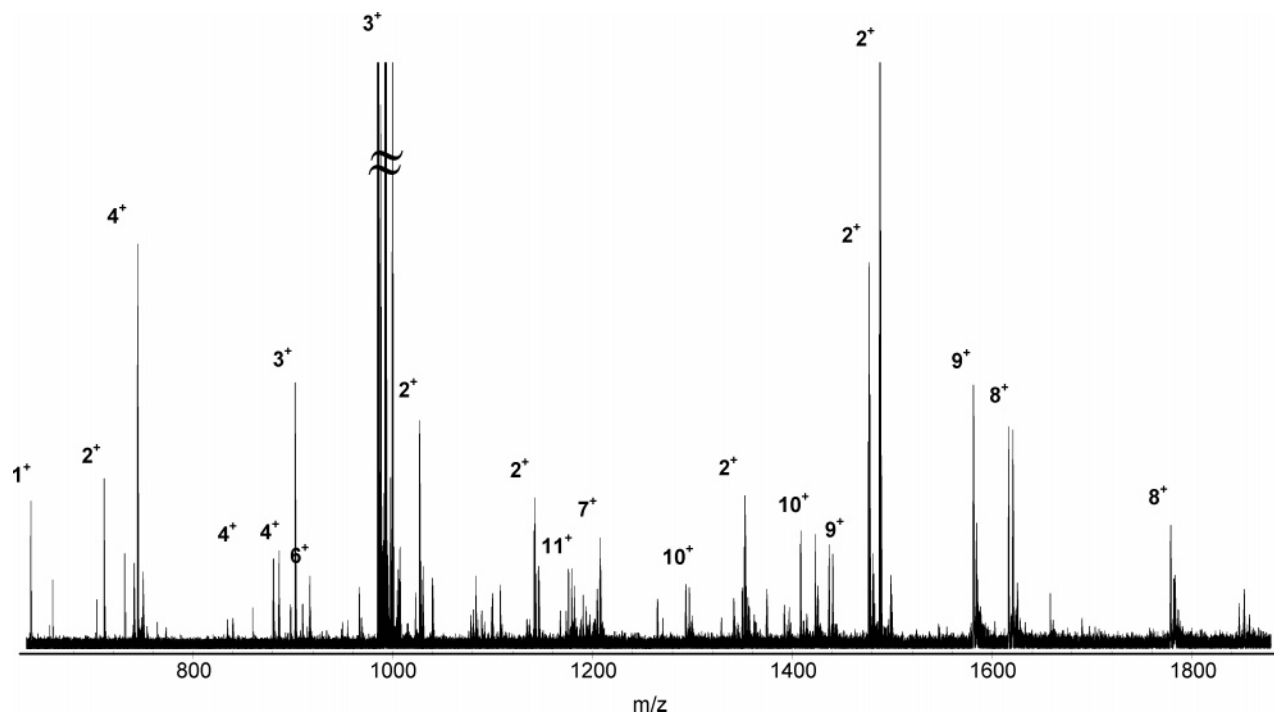


Figure 1. NanoESI FT-ICR mass spectrum of the crude venom of *A. irregularis*. The charge states depicted for the most intense peaks show a broad mass distribution.

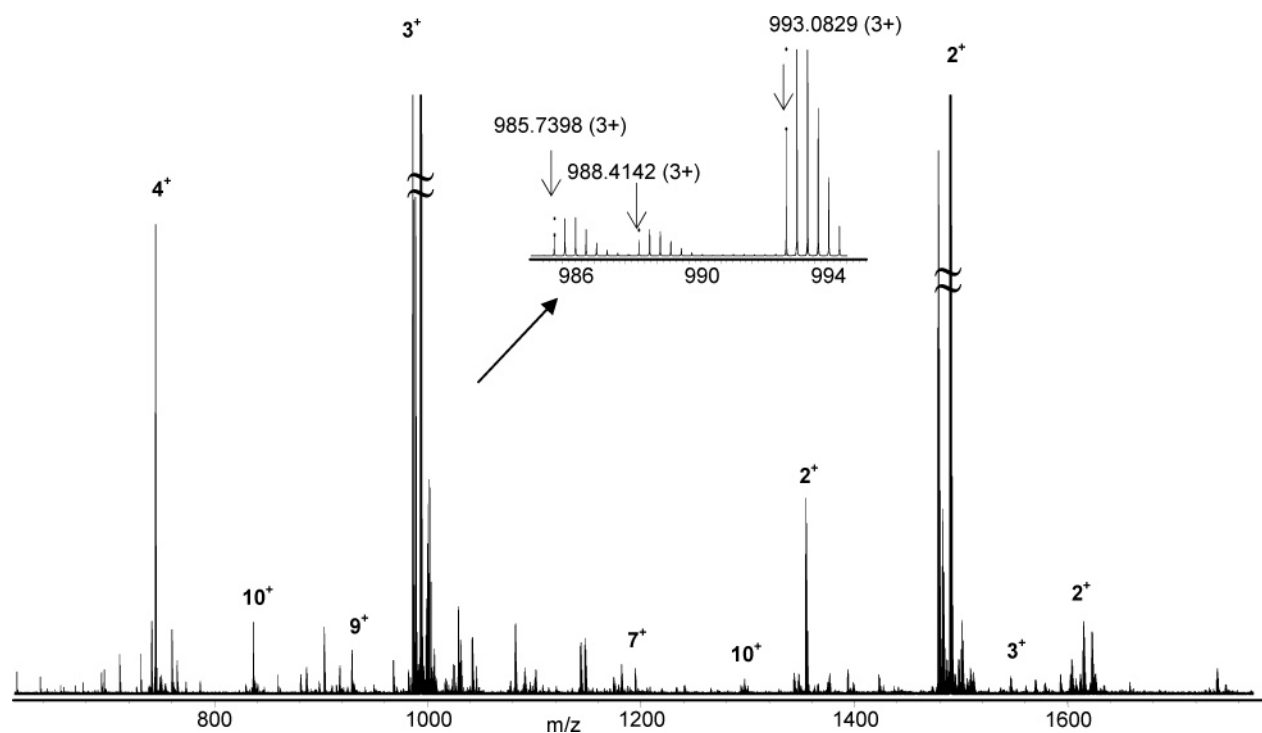


Figure 2. NanoESI FT-ICR mass spectrum of the crude venom of *A. irregularis* after reduction of all disulfide bonds. Zoom on ions between m/z 985 and 993, which will be further fragmented, shows the interest of high resolution and good accuracy for such complex mixtures.

61-amino acid-long propeptide, where the ORF ends after the invariant SRTX-Trp₂₁ by 13 amino acid residues (Figure 3A). Also, we can notice the presence of three dibasic pairs of residues, which could constitute possible cleavage sites for trypsin-like processing enzymes that are often involved in the posttranslation maturation processes of numerous precursors such as pro-enkephalin to enkephalin³² or preproendothelin to big-endothelin (big-ET).^{33,34} Surprisingly, this precursor as well as the other one

we cloned (not shown) displays a different organization as compared to the precursors of SRTXs from *A. engaddensis*²¹ or *A. microlepidota microlepidota*.²² Indeed, when several tandem copies of mature SRTXs are encoded per molecule of the so-called “rosary-type” precursors in the two previous studied *Atractaspididae* snakes, our results clearly establish that in the case of *A. irregularis* only one copy of mature sarafotoxin emerges from each precursor. This constitutes the first description of two structurally

A) 5' gccacctgctttgctcggcccgga ATG GCC CTC CTC CCC CGA CTG GCG GCC
M A L L P R L A A
GGC GGG CTG CTG CTG CTG GGC CTC GCC GCC CTC GAC GGG AAG CCG GCG
G G L L L L L A L A A L D G ↓ K P A
CCG CCG AAG CTG CTG CAG AAG CTG ATG GAC GGG GGG CAG CGG CGC TCC GAG
P P K L L Q K L M D G G Q R R S E
GAC CAG GCA GCA GCA GGA CGC ATC ATC GAC TAC GAG GAC GGA GAT GAA CCG
D Q A A A G R I I D Y E D G D E P
GTC GCT GTC TCG GTG GGA GAC ACG AAA CAG GCC GCG AGA GCC CTC TCA CCG
V A V S V G D T K Q A A R A L S P
CTG CGC AAG CCG CAG CCA CTT TGC TCC TGT ACA GAC ATG TCG GAT CTA GAG
L R K P Q P L *C₁ S C T D M S D L E*
TGC ATG AAT TTC TGC CAT AAG GAC GTC ATC TGG ATA AAC CGG AAC CGG AAA
C M N F C H K D V I W₂₁ I₂₂ N₂₃ R₂₄ N₂₅ R K
CCA TCG CCA ATC CAG AGT TCC tag gctagctaggaagacatccatctc 3'
P S P I Q S S stop

B) MALLPRLAAGGLLLLLLALAADG/KPAPPKLLQKLMDDGGQRRSEDQAAAGRIIDYEDGDE **SRTX-i1**
-----M-----/----- **SRTX-i2**
PVAVSVGDTKQAARALSPLRKPQPL*CSCTDMSDLECMNFCHKDVIWINRNRKPSPIQSS*
-----A-----RL--M-V-----

Figure 3. (A) Nucleotide and deduced amino acid sequences of the cDNA encoding *A. irregularis* SRTX isoform 1. The amino acid sequence is given in single letters. Positions of the PCR primers are underlined. The arrow indicates the probable cleavage site of the leader sequence. The mature 25-amino acid-long SRTX isoform is in italics. Predicted double-basic processing sites are underlined. (B) Deduced amino acid sequences of the two cloned SRTX precursors. The name of the isoforms are indicated at right. “/” indicates the putative cleavage site of the signal peptide. The classical sequence signature of sarafotoxins is indicated in italics. Identical amino acid residues are represented by “—”.

distinct precursor organizations that both encode for toxin isoforms belonging to the same superfamily of vasoactive compounds.

The highly similar amino acid sequences deduced of the two SRTX precursors cloned in this study are shown in Figure 3B and reveal two different SRTX isoforms from *A. irregularis* venom named SRTX-i1 and SRTX-i2. However, despite their distinct overall organization, these new SRTX precursors share several common features with the rosary-type precursors. Thus, in both types of precursor, the predicted signal peptides display 95% of the sequence identity. Also, the prosequence located between the predicted signal peptides and the unique (this study) or the first SRTX sequence (rosary-type precursors of *A. engaddensis*) are similar both in length (61 versus 69 amino acids) and in sequence

since they display 72% of sequence identity, homology, or both (not shown). Together these observations strongly suggest that the genes encoding these two types of precursor have followed a different evolutionary pathway. A plausible complete duplication of a common ancestral gene occurs in the case of *A. irregularis*, although the rosary-type precursor organization most probably reflects a combined duplication/mutation event of the SRTX-encoding exon only. That hypothesis is further supported by the structural characterization of an *A. engaddensis* exon encoding SRTX-c.³⁵ Indeed, the latter contains the sequence of a mature 21-amino acid-long SRTX preceded by a 16-amino acid-long propeptide that appears conserved in both types of SRTX precursors, suggesting that such a SRTX-encoding exon might be present in the corresponding genes.

Comparison between Cloning Results and Nanoelectrospray-FTMS Data. The exact mass for each peptide, deduced from cloning results and beginning with the amino acid sequence

(32) Gubler U.; Seeburg, P.; Hoffman, B. J.; Gage, P. L.; Udenfrirnd, S. *Nature* **1982**, 295, 206–208.

(33) Yanagisawa, M.; Kurihara, H.; Kimura, S.; Tomobe, Y.; Kobayashi, M.; Mitsui, Y.; Yazaki, Y.; Goto, K.; Masaki, T. *Nature* **1988**, 332, 411–415.

(34) Oppenrth, T. J.; Wu-Wong, J. R.; Shiosaki, K. *FASEB J.* **1992**, 6, 2653–2659.

(35) Takasaki, C.; Itoh, Y.; Onda, H.; Fujino, M. *Biochem. Biophys. Res. Commun.* **1992**, 189, 1527–1533.

Table 2. Primary Sequences of the Three New Families of Sarafotoxins (SRTX-i1, -i2, and -i3) Deduced Both from Cloning Results (SRTX-i1 and -i2) and from de Novo Sequencing by FT-ICR Mass Spectrometry (SRTX-i3)^a

measured oxidized M_r	sequence of the new sarafotoxin	no. of amino acids	theoretical oxidized M_r	ΔM_r , ppm	identification
1833.5666	CSCTD MSDLE CMNFC	15	1833.5595	3.8	truncated SRTX-i1
2076.6872	CSCTD MSDLE CMNFC HKD	18	2076.6814	2.7	truncated SRTX-i1
2175.7575	CSCTD MSDLE CMNFC HKDV	19	2175.7498	3.5	truncated SRTX-i1
2288.8368	CSCTD MSDLE CMNFC HKDVI	20	2288.8339	1.3	truncated SRTX-i1
2474.9249	CSCTD MSDLE CMNFC HKDVI W	21	2474.9132	4.7	truncated SRTX-i1
2702.0493	CSCTD MSDLE CMNFC HKDVI WIN	23	2702.0402	3.4	truncated SRTX-i1
2972.1857	CSCTD MSDLE CMNFC HKDVI WINRN	25	2972.1842	0.5	SRTX-i1
2988.1902	CSCTD MSDLE CMNFC HKDVI WINRN ^b	25	2988.1791	3.7	SRTX-i1 ^b
3004.1853	CSCTD MSDLE CMNFC HKDVI WINRN ^c	25	3004.1741	3.7	SRTX-i1 ^c
3353.4441	CSCTD MSDLE CMNFC HKDVI WINRN RKP	28	3353.4331	3.3	elongated SRTX-i1
3537.5179	CSCTD MSDLE CMNFC HKDVI WINRN RKPSP	30	3537.5179	3.3	elongated SRTX-i1
1935.6807	CSCAD MSDLE CMNFC RL	17	1935.6752	2.8	truncated SRTX-i2
2050.7077	CSCAD MSDLE CMNFC RLD	18	2050.7021	2.7	truncated SRTX-i2
2280.8160	CSCAD MSDLE CMNFC RLDVM	20	2280.8110	2.2	truncated SRTX-i2
2296.8174	CSCAD MSDLE CMNFC RLDVM ^b	20	2296.8060	4.9	truncated SRTX-i2 ^b
2680.0071	CSCAD MSDLE CMNFC RLDVM WVN	23	2680.0017	2.0	truncated SRTX-i2
2950.1462	CSCAD MSDLE CMNFC RLDVM WVN RN	25	2950.1457	1.7	SRTX-i2
3515.4929	CSCAD MSDLE CMNFC RLDVM WVN RN RKPSP	30	3515.4794	3.8	elongated SRTX-i2
2688.0291	CSCTD MSDLE CMNFC HKDVI WVN	23	2688.0245	1.7	truncated SRTX-i3
2958.1769	CSCTD MSDLE CMNFC HKDVI WVN RN	25	2958.1686	2.8	SRTX-i3
2966.1492	CSCTD MSDLE CMNFC HKDVI WVN RN ^b	25	2966.1407	2.9	SRTX-i3 ^b
2982.1409	CSCTD MSDLE CMNFC HKDVI WVN RN ^c	25	2982.1356	1.8	SRTX-i3 ^c

^a All SRTX-i1 and SRTX-i2 isoforms derive from cDNA clones 1 and 2, while SRTX-i3 results from de novo sequencing by FT-ICR mass spectrometry only. ΔM_r in ppm corresponds to the difference between the monoisotopic experimental masses of the oxidized experimental species and the theoretical ones. ^b Corresponds to a monooxidized form of the peptide (+15.9949). ^c Corresponds to a doubly oxidized form of the peptide (+31.9898).

“CSC” that is typical of SRTXs, was calculated and compared to the data of Table 1. It appeared immediately that many masses determined from the crude venom of *A. irregularis* (Table 1) correspond to peptides beginning with the sequence CSC and, furthermore, that they could all be deduced with good mass accuracy (0.5–4.7 ppm) from the clones 1 and 2 (Figure 3B). We can also notice that the mRNA molecules that have been cloned correspond to the most abundant peptides in our nanoelectrospray spectrum. The results are summarized in Table 2.

One particularly interesting observation is that the propolypeptide encoded by precursor 1 or 2 appears processed, not in one unique compound but rather in different length molecules ranging from 15 for the shorter to 30 amino acids for the longer. Nanoelectrospray being a very gentle ionization technique, it is unlikely that the existence of these variable-length compounds results from a fragmentation of the peptide ions in the gas phase, but more probably reflects the natural molecular diversity of the venom in sarafotoxin-like peptides. This has already been observed in the venom of *A. microlepidota microlepidota*²² and can be explained either by a different level of degradation in solution of an initial 30-amino acid SRTX isoform or by a different state of maturation of the sarafotoxin propolypeptides.

De Novo Sequencing by Mass Spectrometry within the Crude Venom. To confirm the sequence of the compounds observed for each clone, MS/MS experiments were undertaken on the reduced molecular species that had the highest intensities within the complex mixture. The data obtained for the triply charged ions of m/z 993.030, 985.740, and 988.414 are reported in Tables 3–5. These ions correspond respectively to molecular masses of 2976.2238, 2954.1864, and 2962.2160 Da after the reduction of the disulfide bonds. Although these masses are quite

Table 3. Assignment of y and b Ions Obtained in the SORI-CAD Spectrum^a of the Triply Charged Ion m/z 993.030 Corresponding to a Reduced Molecular Mass of 2976.2238 Da and to the Following Sequence: CSCTD MSDLE CMNFC HKDVI WINRN

ion type	exptl mass	theor mass	error (ppm)	ion type	exptl mass	theor mass	error ^b (ppm)
y ₂	289.1625	289.1613	4.15	b ₅	510.1333	510.1323	1.96
y ₃	403.2053	403.2043	2.48	b ₆	641.1736	641.1728	1.25
y ₄	516.2892	516.2883	1.74	b ₇	728.2048	728.2035	1.79
y ₅	702.3694	702.3676	2.56	b ₈	843.2317	843.2328	1.30
y ₆	815.4549	815.4517	3.92	b ₉	956.3158	956.3128	3.14
y ₇	914.5205	914.5201	0.44	b ₁₂	1319.4081	1319.4134	4.02
y ₈	1029.5531	1029.5471	5.83	b ₁₃	1433.4510	1433.4606	6.70
y ₉	1157.6451	1157.6420	2.68	b ₁₄	1580.5194	1580.5246	3.29
y ₁₀	1294.7004	1294.7009	0.39	b ₁₆	1820.5875	1820.5777	5.38
y ₁₁	1397.7099	1397.7101	0.14	b ₁₈	2063.7094	2063.7173	3.83
y ₁₂	1544.7764	1544.7785	1.36	b ₁₉	2162.7778	2162.7718	2.77
y ₁₃	1658.8206	1658.8215	0.54	b ₂₀	2275.8619	2275.8533	3.78
y ₁₅	1892.8745	1892.8711	1.80	b ₂₁	2461.9412	2461.9457	1.83
y ₁₇	2135.0049	2134.9978	3.33	b ₂₂	2575.0253	2575.0302	1.90
y ₁₈	2250.0343	2250.0247	4.27	b ₂₃	2689.0682	2689.0783	3.76
y ₁₉	2337.0625	2337.0568	2.44	b ₂₄	2845.1693	2845.1709	0.56
y ₂₀	2468.1064	2468.0972	3.73				
y ₂₁	2583.1143	2583.1242	3.83				
y ₂₃	2787.1932	2787.1810	4.38				

^a Conditions: 100 scans, 500 Hz off-resonance, 5 pulses of Xe up to 10⁻⁷ mbar, 400 μ s irradiation time, and $V_{pp} = 4$ V. ^b The error in ppm corresponds to the difference between theoretical data and experimental ones (taking into account the mass of the electron).

high for de novo sequencing, a large part of the primary sequence can be easily retrieved in each case.

For m/z 993.030, which is the most intense of the three ions (zoom in Figure 2), a high coverage (87.5%) of the sequence is obtained (21/24 bonds cleaved) with nice series of both b and y ions.³⁶ The mass accuracy for all fragments lies in the range 0.5–7

Table 4. Assignment of y and b Ions Obtained in the SORI-CAD Spectrum^a of the Triply Charged Ion m/z 985.740 Corresponding to a Reduced Molecular Mass of 2954.1864 Da and to the Following Sequence: CSCAD MSDLE CMNFC RLDVM WVNRN

ion type	exptl mass	theor mass	error ^b (ppm)
y ₂	289.1613	289.1627	4.84
y ₃	403.2043	403.2057	3.47
y ₄	502.2727	502.2741	2.79
y ₅	688.352	688.3533	1.89
y ₆	801.436	801.437	1.25
y ₇	900.5045	900.5059	1.55
y ₁₇	2120.9821	2120.9818	0.14
y ₁₉	2323.0411	2323.0242	7.27
y ₂₀	2454.0816	2454.1033	8.84
b ₄	395.1054	395.1027	6.83
b ₁₈	2063.7094	2063.6999	4.60
b ₁₉	2162.7778	2162.7915	6.33
b ₂₀	2275.8619	2275.8533	3.78
b ₂₁	2461.9412	2461.9275	5.56
b ₂₂	2561.0096	2561.0271	6.83
b ₂₄	2831.1537	2831.1396	4.98

^a Conditions: 150 scans, 500 Hz off resonance, 5 pulses of Xe up to 10^{-7} mbar, 400 μ s irradiation time, and $V_{pp} = 4$ V. ^b The error in ppm corresponds to the difference between theoretical data and experimental ones (taking into account the mass of the electron).

Table 5. Assignment of y and b Ions Obtained in the SORI-CAD Spectrum^a of the Triply Charged Ion M/z 988.414 Corresponding to a Reduced Molecular Mass of 2962.216 Da and to the Following Sequence: CSCTD MSDLE CMNFC HKDVI WVNRN

ion type	exptl mass	theor mass	error ^b (ppm)
y ₂	289.1613	289.1627	4.84
y ₃	403.2043	403.2057	3.47
y ₄	502.2727	502.2741	2.79
y ₅	688.352	688.3533	1.89
y ₆	819.3925	819.3941	1.95
y ₇	918.4609	918.4655	5.01
y ₈	1033.4878	1033.4886	0.77
y ₁₉	2345.0288	2345.0134	6.57
b ₁₈	2037.7302	2037.734	1.86
b ₁₉	2136.7986	2136.7948	1.78
b ₂₀	2267.8391	2267.8366	1.10
b ₂₁	2453.9184	2453.8958	9.21
b ₂₂	2552.9868	2552.9705	6.38
b ₂₃	2667.0297	2667.0173	4.65
b ₂₄	2823.1308	2823.1228	2.83

^a Conditions: 150 scans, 500 Hz off resonance, 5 pulses of Xe up to 10^{-7} mbar, 400 μ s irradiation time, $V_{pp} = 4$ V. ^b The error in ppm corresponds to the difference between theoretical data and experimental ones (taking into account the mass of the electron).

ppm, which is good for MS/MS experiments. The nice extent of fragmentation obtained can be explained both by the use of xenon as the collision gas, which increases the fragmentation efficiency by a factor of 6 in comparison to the use of argon, and by the fact the initial ion bears three protons, which increases the Coulombic repulsion energy of the system. Moreover, the intensity of the selected ions before fragmentation is high, which yields a good signal/noise ratio for the fragments. The major fragmentation observed is formation of b₂₄, which can be explained by the

location of the arginine residue at the C-terminus of the sequence that retains the mobile proton during the fragmentation. The full coverage of the major species observed for clone 1 (25 amino acids long) together with the number of disulfide bonds, and the high accuracy of the molecular mass, give high confidence for all structures derived directly from this clone, i.e., structures of 18, 19, 20, 23, 25, 28, and 30 amino acids. Moreover, the characteristics of these seven species constitute a new family of sarafotoxin-related peptides that will be called, for example, SRTX-i1₁₈ in the case of the isoform of 18 amino acids derived from clone 1. Within this new family of compounds, the most abundant molecular species is 25 amino acids long with four additional C-terminal residues: "V/I₂₂NRN₂₅" as compared to SRTXs from *A. engadensis*.^{20,21} It is noteworthy that the C-terminus extension of SRTX-i1₂₅ is unrelated to that of long-SRTX from *A. microlepidota* microlepidota: V/I₂₂NRN₂₅ versus D/E₂₂Q/N₂₃P/V₂₄. Even though these isoforms share a common structural pattern with other sarafotoxins, we can nevertheless notice a diversification in the amino acid composition since three new amino acids are found in positions that were highly conserved previously: (i) a threonine residue at position 4, (ii) a leucine residue at position 9, and (iii) a lysine at position 17 (Figure 4).

The tandem mass spectrum of the ion at m/z 985.740 confirms the presence of sarafotoxins related to the complementary DNA (cDNA) clone 2 since the C-terminus tag "V₁₉MWVNRN₂₅" is retrieved easily. The absence of ions arising from a N-terminal fragmentation, confirming the presence of the "CSCAD" moiety, can be explained by the location of the charge in C-terminus region as shown by the b and y ions obtained. As for SRTX-i1, this new family of sarafotoxins (SRTX-i2) contains several members with various lengths (17–25 amino acids) and an elongated one (30 amino acids) that, again, raises the question of the maturation of a long form of the propeptide. As found for SRTX-i1, SRTX-i2 increases the diversity of the sarafotoxins family since new amino acids are found in the backbone. Even though alanine and leucine, in positions 4 and 17, respectively, have previously been observed in other *Atractaspis* samples, the location of a methionine in position 20, as well as the additional C-terminal amino acids, is completely novel.

The MS/MS spectrum of the ion at m/z 988.414 is also very informative since it demonstrates that the difference of 14.009 Da observed between the molecular masses of the oxidized species at 2972.1857 and 2958.1769 Da can be explained simply by the substitution of the isoleucine in position 22 by a valine. This is clearly shown when calculating b₂₂–b₂₁ ($\Delta = 99.0684$) or y₄–y₃ ($\Delta = 99.0684$), which corresponds exactly to the mass of a valine (99.06841 Da). The theoretical mass difference between Ile and Val is 14.016 Da (as is the difference between S and T, D and E, N and Q, or A and G). This new isoform of SRTX-i (named SRTX-i3) has been retrieved with at least two lengths, 23 and 25 amino acids, since the shorter forms are common with those of SRTX-i1. Again, the 25-amino acid-long isoform is the most abundant. The characterization of that third family of compounds relies only on mass spectrometric results since its amino acid sequence does not match with any cloning results. As for SRTX-i1₂₅, SRTX-i3₂₅ is also retrieved under both mono- and dioxidized forms. Finally, SRTX-i1 and SRTX-i3 have very similar structures that differ by only one amino acid: a valine (SRTX-i3) or an isoleucine (SRTX-

(36) Roepstorff, P.; Fohlman, J. *Biomed. Mass Spectrom.* **1984**, *11*, 601.

	1			5				10				15				21				24						
A)	Cys	Ser	Cys	Lys	Asp	Met	Thr	Asp	Lys	Glu	Cys	Leu	Asn	Phe	Cys	His	Gln	Asp	Val	Ile	Trp				SRTX-a	
	Cys	Ser	Cys	Lys	Asp	Met	Thr	Asp	Lys	Glu	Cys	Leu	Tyr	Phe	Cys	His	Gln	Asp	Val	Ile	Trp				SRTX-b	
	Cys	Thr	Cys	Asn	Asp	Met	Thr	Asp	Glu	Glu	Cys	Leu	Asn	Phe	Cys	His	Gln	Asp	Val	Ile	Trp				SRTX-c	
	Cys	Ser	Cys	Ala	Asp	Met	Thr	Asp	Lys	Glu	Cys	Leu	Tyr	Phe	Cys	His	Gln	Asp	Val	Ile	Trp				Btx	
B)	Cys	Thr	Cys	Phe	Thr	Tyr	Lys	Asp	Lys	Glu	Cys	Val	Tyr	Tyr	Cys	His	Leu	Asp	Ile	Ile	Trp				ET-3	
	Cys	Ser	Cys	Ser	Ser	Leu	Met	Asp	Lys	Glu	Cys	Val	Tyr	Phe	Cys	His	Leu	Asp	Ile	Ile	Trp				ET-1	
	Cys	Ser	Cys	Ser	Ser	Trp	Leu	Asp	Lys	Glu	Cys	Val	Tyr	Phe	Cys	His	Leu	Asp	Ile	Ile	Trp				ET-2	
	Cys	Ser	Cys	Asn	Ser	Trp	Leu	Asp	Lys	Glu	Cys	Val	Tyr	Phe	Cys	His	Leu	Asp	Ile	Ile	Trp				VIC	
	Cys	Ser	Cys	Ala	Thr	Phe	Leu	Asp	Lys	Glu	Cys	Val	Tyr	Phe	Cys	His	Leu	Asp	Ile	Ile	Trp				ET-trout	
C)	Cys	Ser	Cys	Lys	Asp	Met	Ser	Asp	Lys	Glu	Cys	Leu	Asn	Phe	Cys	His	Gln	Asp	Val	Ile	Trp				SRTX-a1	
	Cys	Ser	Cys	Lys	Asp	Met	Ser	Asp	Lys	Glu	Cys	Leu	Tyr	Phe	Cys	His	Gln	Asp	Val	Ile	Trp				SRTX-b1	
	Cys	Thr	Cys	Lys	Asp	Met	Thr	Asp	Lys	Glu	Cys	Leu	Tyr	Phe	Cys	His	Gln	Gly	Ile	Ile	Trp				SRTX-e	
D)	Cys	Ser	Cys	Asn	Asp	Ile	Asn	Asp	Lys	Glu	Cys	Met	Tyr	Phe	Cys	His	Gln	Asp	Val	Ile	Trp	<i>Asp⁵</i>	<i>Glu²</i>	<i>Pro³</i>	SRTX-m	
	Cys	Ser	Cys	Asn	Asp	Met	Asn	Asp	Lys	Glu	Cys	Met	Tyr	Phe	Cys	His	Gln	Asp	Val	Ile	Trp	<i>Asp⁵</i>	<i>Glu²</i>	<i>Pro³</i>	SRTX-m1	
	Cys	Ser	Cys	Asn	Asp	Ile	Asn	Asp	Lys	Glu	Cys	Met	Tyr	Phe	Cys	His	Gln	Asp	Ile	Ile	Trp	<i>Asp⁵</i>	<i>Glu²</i>	<i>Pro³</i>	SRTX-m2	
	Cys	Ser	Cys	Asn	Asp	Met	Asn	Asp	Lys	Glu	Cys	Val	Tyr	Phe	Cys	His	Leu	Asp	Ile	Ile	Trp	<i>Asp⁵</i>	<i>Glu²</i>	<i>Pro³</i>	SRTX-m3	
	Cys	Ser	Cys	Asn	Asn	Met	Ser	Asp	Lys	Glu	Cys	Leu	Asn	Phe	Cys	Asn	Leu	Asp	Ile	Ile	Trp	<i>Glu¹</i>	<i>Asn⁴</i>	<i>Val¹</i>	SRTX-m4	
	Cys	Ser	Cys	Asn	Asp	Met	Asn	Asp	Lys	Glu	Cys	Met	Tyr	Phe	Cys	His	Gln	Asp	Ile	Ile	Trp	<i>Asp⁵</i>	<i>Glu²</i>	<i>Pro³</i>	SRTX-m5	
E)	Cys	Ser	Cys	Thr	Asp	Met	Ser	Asp	Leu	Glu	Cys	Met	Asn	Phe	Cys	His	Lys	Asp	Val	Ile	Trp	<i>Ile¹</i>	<i>Asn⁴</i>	<i>Arg²</i>	<i>Asn²</i>	SRTX-i1 ₂₅
	Cys	Ser	Cys	Ala	Asp	Met	Ser	Asp	Leu	Glu	Cys	Met	Asn	Phe	Cys	Arg	Leu	Asp	Val	Met	Trp	<i>Val¹</i>	<i>Asn⁴</i>	<i>Arg²</i>	<i>Asn²</i>	SRTX-i2 ₂₅
	Cys	Ser	Cys	Thr	Asp	Met	Ser	Asp	Leu	Glu	Cys	Met	Asn	Phe	Cys	His	Lys	Asp	Val	Ile	Trp	<i>Val¹</i>	<i>Asn⁴</i>	<i>Arg²</i>	<i>Asn²</i>	SRTX-i3 ₂₅
F)	Cys	Ser ¹⁸ Thr ³	Cys	Asn ⁴ Lys ⁵ Ser ²	Asp ¹⁵ Ser ² Trp ²	Met ¹⁴ Ile ² Trp ²	Ser ⁶ Thr ² Asn ⁵	Asp	Lys ¹⁷ Leu ³ Glu ¹	Glu	Cys	Leu ⁸ Met ⁷ Val ⁶	Tyr ¹⁴ Asn ⁷	Phe ²⁰ Tyr ¹	Cys	His ¹⁹ Arg ¹ Asn ¹	Gln ¹¹ Leu ⁸ Lys ²	Asp ²⁰ Gly ¹	Val ¹¹ Ile ¹⁰	Ile ²⁰ Met ¹	Trp	Asp ⁵ Val ² Ile ¹	Glu ² Asn ⁴	Pro ³ Arg ² Val ¹	Asn ²	

Figure 4. Sequence alignments of various SRTXs and ETs. The numbering is indicated in the first line. (A) SRTX isoforms previously isolated from the venom of *A. engaddensis* (SRTX-a, -b, and -c) and from *Atractaspis bibroni* (Btx). (B) Endothelins from mammals, VIC peptide (vasoactive intestinal peptide), and ET from trout. (C) Deduced SRTX sequences from *A. engaddensis* precursors. (D) De novo characterized SRTX sequences from *A. microlepidota microlepidota* precursor. Additional residues are in italics. (E) De novo characterized SRTX sequences from *A. irregularis*. (F) Amino acid variability as deduced from previously described sequences. Residues in boldface type are invariants. The frequencies of the different amino acid residues within the 21 analyzed sequences are indicated as upper signs.

i1) at position 22 following the conserved Trp₂₁ (Figure 4). Note that the mass accuracy provided by FTMS is required: (i) for confidence when comparing the molecular masses determined experimentally with the theoretical molecular masses deduced from cloning; (ii) for the determination of the primary sequences in de novo sequencing experiments when some fragment ions are missing in the MS/MS spectra.

It is noteworthy that none of the molecular masses corresponding to the SRTX-i3 family are derived from the two precursors cloned. One explanation could be a failure to clone the corresponding precursor, either due to a particularly low abundance of the latter or because of an inadequacy of our set of primers. Both hypotheses are unlikely, first because the relative abundance of SRTX-i3 peptides is as high as those of SRTX-i1 and -i2 isoforms (Table 1), and second because our set of PCR primers has proven their efficacy in isolation of cDNA clones 1 and 2 encoding SRTX-i1 and -i2, respectively. A more plausible explanation would be that SRTX-i3 isopeptides derive from an edited mRNA, as has been proposed in the case of α -bungarotoxin isoforms present in the venom of the snake *Bungarus multicinctus*.³⁷ Even if the mechanism of posttranscriptional modification of RNA termed "RNA editing" in snake cells is unknown, it corresponds in mammalian cells to a deamination of cytosine and adenine to form uracil and inosine that is performed by the APOBEC-1 and DRADA deaminases, respectively.³⁸ To date, four classes of RNA editing have been observed in cells: C \rightarrow U, A \rightarrow G, U \rightarrow C, and U \rightarrow A, leading to an increase of the proteic polymorphism within a particular tissue or type of cells. Thus, the single amino acid difference that does exist between SRTX-i1 and SRTX-i3, Ile₂₂ versus Val₂₂ (Table 2 and Figure 4) could be

easily explained by a adenine (A) \rightarrow guanine (G) nucleotide editing of the first base within the Ile codons, AU/UCA (Ile) to generate Val encoding codons GU/UCA. However, only an analysis of the genomic DNA of SRTX from *A. irregularis*, would allow us to confirm or deny that hypothesis.

CONCLUSION

In this work, it has been shown that a combined use of high-resolution mass spectrometry (FTMS) and molecular cloning can be a powerful tool to characterize new active molecules directly within very complex mixtures, such as crude venoms, without chromatographic separation.

The venom chosen is a unique sample from *A. irregularis*, a species never previously studied at the molecular level. This snake belongs to the *Atractaspidae* family, which is known to produce highly toxic venoms that contain endothelin-like peptides called sarafotoxins (SRTXs). Recently, these vasoactive peptides were found to be 21 (*A. engaddensis*) or 24 amino acids long (*A. microlepidota microlepidota*) with two conserved disulfide bridges and were produced from original precursors. Our study reveals the presence of three new families of sarafotoxins (SRTXi1–3), some of them being longer by up to 9 amino acids than the 21-amino acid-long SRTXs. The high performances of FTMS (high mass accuracy and resolution) were required for de novo sequencing of each family and allowed great confidence in all primary sequences found.

Surprisingly, the precursors encoding the new sarafotoxin isoforms 1 and 2 in *A. irregularis* display a different overall organization as compared to those encoding SRTXs from *A. engaddensis* and *A. microlepidota microlepidota*. Indeed, when several tandem copies of mature SRTXs are encoded per molecule of the so-called rosary-type precursor in the two previous studied

(37) Liu, L. F.; Chang, C. C.; Liao, M. Y.; Kuo, K. W. *Nucleic Acids Res.* **1998**, *26*, 5624–5629.

(38) Ashkenas, J. *Am. J. Hum. Genet.* **1997**, *60*, 278–283.

Atractaspididae snakes, our results clearly establish that in the case of *A. irregularis* only one copy of mature sarafotoxin can emerge from each precursor. This constitutes the first description of two structurally distinct precursor organizations that both encode for toxin isoforms belonging to the same superfamily of vasoactive compounds, suggesting a different evolutionary pathway.

Together our results further illustrate the great diversity of that family of compounds from a structural point of view, and

SRTX-i1₂₅ as deduced from mRNA clone 1 is actually chemically synthesized, to initiate the study of its three-dimensional structure by NMR, and also to establish its biological properties and toxicity.

ACKNOWLEDGMENT

We thank Pr. I. Ineich (Museum National d'Histoire Naturelle, Paris) for providing the specimen of *Atractaspis irregularis*.

Received for review April 5, 2005. Accepted July 6, 2005.

AC050575K

Fourier transform mass spectrometry: A powerful tool for toxin analysis

Loïc Quinton^a, Jean-Pierre Le Caër^a, Joëlle Vinh^b, Nicolas Gilles^c,
Julia Chamot-Rooke^{a,*}

^a Laboratoire des Mécanismes Réactionnels, UMR 7651 CNRS, Ecole Polytechnique, F-91128 Palaiseau, France

^b Ecole Supérieure de Physique et de Chimie Industrielles de la ville de Paris, 10 rue Vauquelin, 75231 Paris Cedex, France

^c Département d'Ingénierie et d'Etudes des Protéines, Commissariat à l'Energie Atomique, Centre d'Etudes de Saclay, 91191 Gif-sur-Yvette Cedex, France

Available online 30 March 2006

Abstract

The crude venom of *Conus virgo* was analyzed by Fourier transform mass spectrometry (FTMS) using both nano-electrospray ionization and MALDI. The analyses were performed directly on the crude venom, without chromatographic separation. The mass fingerprinting of the venom yielded 64 distinct molecular masses in the range 500–4500 Da with two major components at 1328.5142 and 1358.5592 Da. To facilitate the de novo sequencing of these compounds, the disulfide bonds of all components were reduced for the whole venom. The mass accuracy, resolution and sensitivity provided by FTMS were necessary to complete the sequencing of the two new peptides named ViVA and ViVB, that turned out to be conotoxins belonging to the T-superfamily, with the disulfide framework V. The peptides shared 80% similarity and as often observed for this class of compound, they were highly post-translationally modified: amidated C-terminus, pyroglutamic acid residue at the N-terminus and two disulfide bonds. Complementary online nano-LC-nano-ESI-FTMS experiments were undertaken. Among the 130 molecular masses found in the coupling experiments, only 45 were common with those obtained in the direct approach, which means that 21 compounds observed by nano-ESI-FTMS were not detected. This clearly shows that some discriminations against some classes of compounds occur when a chromatographic step is used before mass spectrometry.

© 2006 Elsevier Ltd. All rights reserved.

Keywords: Fourier transform mass spectrometry; Nano-ESI; *Conus virgo*; Conotoxin; De novo sequencing; Post-translational modifications

1. Introduction

Since the development of electrospray (ESI) and matrix assisted laser desorption ionization (MALDI), two soft ionization sources allowing mass spectrometry (MS) to be used by biologists, the use of this powerful analytical technique for analysis of peptides, proteins, carbohydrates or other biomolecules has become essential (Yates, 2000). Moreover, the easy coupling between nano-liquid

chromatography and nano-electrospray (nano-flow versions of LC and ESI), made mass spectrometry the method of choice for the analysis of complex mixtures, such as enzymatic digests of proteins. Although the most popular application of mass spectrometry is protein identification (usually after digestion and mass fingerprinting) in databases for genomes sequenced (Liska and Shevchenko, 2003), different structural information can be obtained from an appropriate mass spectrometric analysis: accurate determination of molecular mass, amino acid sequence for peptides (Spengler, 2004), determination of disulfide bonds (Gorman et al., 2002), characterization of post-translational modifications using the recent technique of electron capture dissociation (Cooper et al., 2005; Kelleher et al., 1999),

* Corresponding author. Tel.: +33 1 69 33 34 07; fax: +33 1 69 33 30 41.

E-mail address: jcr@dcmr.polytechnique.fr (J. Chamot-Rooke).

multimeric state of proteins, structural informations for bioorganic compounds. Nevertheless, it is clear that the performance of the mass spectrometer used for the analysis is of prime importance for the quality of the data obtained. Fourier transform mass spectrometry (FTMS) (Marshall and Hendrickson, 2002; Marshall et al., 1998; Marshall, 2004) provides the highest performance in mass resolution and mass accuracy and for that reason is gaining more and more importance in analyzing complex mixtures obtained from protein digests (Li et al., 2001), or simply derived from natural sources such as venoms (Quinton et al., 2005; Wermelinger et al., 2005).

Actually, venoms are highly complex mixtures that can contain several hundred compounds (Kalume et al., 2000), many of them having potent biological activity. Even though most compounds are peptides (which can be highly post-translationally modified such as in the mollusk genus *Conus*) (Olivera, 1997; Olivera et al., 1990), proteins, enzymes (phospholipases) (Bao et al., 2005) or small organic compounds such as polyamines (Tzouros et al., 2004) are often also present in the same venom. The highly diverse biological activities observed for these different molecular species makes venoms a rich source of biologically active compounds. This diversity explains both the high toxicity of some of these venoms and the increasing number of venom screening programs to discover new toxins, often relying on a combination of sample extraction, electrophoresis, reversed phase high performance liquid chromatography (RP-HPLC), mass spectrometry (MS) (Hansson et al., 2004), Edman degradation and/or molecular cloning (Diego-García et al., 2005).

In this study, we report the characterization by Fourier transform mass spectrometric strategies only, of two novel conotoxins belonging to the T-superfamily, in the venom of *Conus virgo*. It will be shown that the high performance of FTMS, and especially the high mass accuracy provided by such instrumentation, were necessary for de novo sequencing (Spengler, 2004; Schlosser and Lehmann, 2002) of these highly modified toxins directly from the crude venom. In order to improve the dynamic range of the system and to evaluate the ion suppression effects due to the use of nano-electrospray, the venom was also analyzed by nano-LC-FTMS. This sensitive coupling allowed the detection of 130 distinct molecular masses in the whole venom.

2. Material and methods

2.1. Mass spectrometer

All experiments were carried out with a 7-T APEX III FT-ICR mass spectrometer (Bruker Daltonics, Bremen, Germany) equipped with a 7-T actively shielded superconducting magnet and an infinity cell (Caravatti and Allemann, 1991). The ionization sources used were nano-electrospray and MALDI.

2.2. Nano-electrospray analysis

Nano-electrospray needles were purchased from Proxeon (Odense, Denmark), filled with 3 μ L of the sample diluted in methanol/water/formic acid 49.9:49.9:0.2 (v/v/v) and the tapered end of the tip subsequently carefully broken under an optical microscope. A stable spray was observed by applying a voltage of about -700 V between the needle (grounded) and the entrance of the glass capillary used for ion transfer. A source temperature of 50°C was used for all nano-electrospray experiments. The estimated flow rate was 20–50 nL/min. Ions were stored in an hexapole guide for 1 s and pulsed into the detection cell through a series of electrostatic lenses. Ions were finally trapped in the cell using SideKick™ and front and back trapping voltages of 0.9 and 0.95 V, respectively. Mass spectra were acquired from m/z 500 to 2000 with 512k data points and monoisotopic peaks were automatically labeled using the software XMASS, version 6.1.4 (Bruker Daltonics). For sustained off resonance irradiation collision activated dissociation (SORI-CAD) (Gauthier et al., 1991) experiments, the ion to be fragmented was isolated by radio frequency (rf) ejection of all unwanted ions using both low-voltage single rf pulses (soft shots) at their resonance frequencies and a chirp excitation covering the region of interest. The fragmentation was performed using an excitation duration of 400 μ s ($V_{pp}=37$ dB) at a frequency 500 Hz higher than the cyclotron frequency of the ion of interest. Xenon was used as the collision gas and introduced several times through a pulsed valve in order to obtain a pressure in the cell up to 10^{-7} mbar. The excitation/detection of all fragment ions was performed after a pumping delay of 4 s. An external calibration using NaI (2 mg/mL in isopropanol/water) was performed daily.

2.3. MALDI analysis

The SCOUT MALDI source (Bruker Daltonics) with pulsed in-source collision gas and in-source accumulation capability was used for this study. This source uses collision gas (argon) introduced during ion formation and accumulation for cooling the ions. Singly protonated peptides were externally produced using a 337 nm nitrogen laser (20 Hz). Ions were stored in the source region and pulsed into the detection cell through a series of electrostatic lenses. Ions were finally trapped in the cell using SideKick™ and front and back trapping voltages of 0.9 and 0.95 V. Mass spectra were acquired from m/z 500 to 4000 with 512k data points and monoisotopic peaks were automatically labeled using the software XMASS 6.1.4 (Bruker Daltonics). For each measurement, 20–30 scans (200–300 laser shots) were summed. External or internal calibration were performed using a peptide calibration standard (Bruker) containing Angiotensin I and II, Substance P, Bombesin, ACTH clip 1–17, ACTH clip 18–39 and Somatostatin 28.

2.4. Nano-LC-FTMS analysis

The crude venom was analyzed by nano-scale reversed phase high performance liquid chromatography (nano-HPLC) using an Agilent 1100 series instruments coupled to the FT-ICR instrument. The venom was concentrated and desalted online on a Zorbax 300SB-C18 (5×0.3 mm) precolumn and the separation of its components was performed on a Zorbax 300SB-C18 (150×0.075 mm). Buffer A was 100% H₂O and 0.1% acid formic (FA) and buffer B was 90% acetonitrile (ACN), 10% H₂O and 0.1% FA. Prior to the chromatographic separation, the sample was loaded on the precolumn for 5 min with a buffer composed of 100% H₂O and 1% FA. A binary gradient of 5–50% solvent B over 45 min was generated and it was followed by a gradient from 50 to 100% solvent B over 12 min. The mobile phase composition was then left at 100% solvent B for 5 min and back to 0% buffer B over 1 min. The sample injection was set up to 2 µL using the Agilent micro sampler and the column flow was reduced by splitting to 300 nL/min. The needles used for the nano-ESI interface were PicoTip™ emitters SilicaTip™ (360 µm (o.d.)×75 µm (i.d.)) with a 10 µm tip opening purchased from New Objective (Woburn, MA, USA). The voltage of the nano-ESI voltage and the heated interface were set to –1300 V and 150 °C, respectively.

2.5. *Conus virgo* venom preparation

Fifteen specimen of *Conus virgo* were collected in the Philippines. Their venom ducts were dissected, then cut into small pieces and extracted with 5% acetic acid. The obtained solutions were pooled and centrifuged. Then, the crude venom (i.e. supernatant) was lyophilized and a small amount of this lyophilized sample was diluted in a large volume of 1% acetic acid (~5 mL) to constitute our stock solution. For each analysis, 100 µL of this stock solution was lyophilized and the residue was dissolved by adding 10 µL of 1% formic acid, to concentrate the sample. For direct infusion by nano-ESI-FTICR, this solution was desalted using ZipTip® C₄ pipette tips (Millipore) and eluted with 10 µL of 49.9:49.9:0.2 methanol/water/formic acid (v/v/v). For nano-HPLC-FTMS, 2 µL of the concentrated solution was directly injected in the chromatographic system without previous purification. For MALDI experiments 2,5-dihydroxy benzoic acid (DHB) (Sigma-Aldrich) was used as the matrix. A 50 mg/mL solution of DHB in water:acetonitrile (70:30 v/v), with the addition of 0.2% trifluoroacetic acid was prepared. One microliter of the concentrated solution of crude venom was vortexed with 1 µL of the DHB solution. 0.25 µL of the resultant mixture was loaded on the MALDI target for each spot.

2.6. Peptides synthesis

Solid phase synthesis of the linear peptides was done with an ABI 430A synthesiser (Perkin-Elmer/Applied

Biosystems Division), using Fmoc-chemistry with DCC/HOAt as the condensation reagent and 0.1 mmol of Rink amide resin. Linear peptide was purified by HPLC on a C18 Vydac analytical column in 0.1% trifluoroacetic acid and eluted with an acetonitrile gradient.

2.7. Reduction procedure

Reduction of the *Conus virgo* crude venom was performed using Tris(2-carboxyethyl)phosphine (TCEP) (Burn et al., 1991). TCEP is an odourless and non-volatile reducing agent that is more stable and effective than dithiothreitol (DTT) (Burmeister Getz et al., 1999). It reduces disulfide bonds rapidly and completely in water at pH 4–5. Ten microliter of the concentrated *Conus virgo* venom solution was incubated with 2 µL of TCEP (10 nmol/µL) at 50 °C for 30 min at pH 4. The course of the reduction was monitored by corresponding shifts in the mass spectrum (the reduction of each S–S bond yields a +2 Da shift for the molecular mass).

3. Results and discussion

3.1. Nano-ESI-FTMS analysis of the crude venom

The nano-ESI-FTMS mass spectrum of the crude venom exhibits about 64 distinct molecular masses in the range 500–4500 Da (Table 1, Fig. 1).

The analysis of the spectrum shows that all species detected are in a large dynamic range and more particularly that two compounds are widely predominant. These compounds (29 and 30 in Table 1) are found in the spectrum with two different charge states (1+ and 2+) and also with several Na⁺ and K⁺ adducts, in spite of the desalting step. The abundances of these two compounds (the measurement is based on the intensity of the monoisotopic peaks for the protonated and cationized species) represent, respectively, 18 and 51% of the total signal (sum of the intensities of all ions). The measured monoisotopic molecular mass are 1328.5135 and 1358.5574 Da (average of the different charge states observed). These masses are in the typical range of molecular masses of conotoxins, which generally correspond to peptides of 10–30 amino acids.

Since these molecular masses have never been observed before in *Conus* venoms, we decided to de novo sequence these two species by mass spectrometry directly from crude venom, without prior HPLC purification.

First of all, to measure these masses with the best mass accuracy possible, MALDI-FTICR experiments were performed with internal calibration. For that purpose, the calibration pepmix from Bruker [mixture of Angiotensin I (MH⁺1296.6853) and II (MH⁺1046.5420), Substance P (MH⁺1347.7361) and Bombesin (MH⁺1619.8230)] was mixed with crude venom and DHB, and spotted on the target. The molecular masses (Mr) obtained for the two

Table 1

Ions in m/z (z =charge state) with their relative intensities (I) in the spectrum and corresponding molecular masses observed in the nano-ESI-FTICR spectrum of the crude venom of *Conus virgo*

m/z	z	% I	Mr	Compound	m/z	z	% I	Mr	Compound
546.4014	1	0.22	545.3941	1	813.3744	2	0.06	1622.7342	36
587.2842	1	0.35	586.2769	2	884.7983	2	0.46	1767.5820	37*
590.4278	1	0.28	589.4205	3	1053.0294	2	0.19	2104.0442	38
634.4547	1	0.32	633.4474	4	776.3031	3	0.17	2325.8875	39
637.3063	1	0.54	636.2990	5	786.3187	3	0.12	2355.9343	40
722.5059	1	0.28	721.4986	6	815.3270	3	0.06	2442.9592	41
739.6071	1	0.20	738.5998	7	886.6868	3	0.17	2657.0386	42
766.5312	1	0.25	765.5239	8	1344.5425	2	0.23	2687.0704	43*
767.4704	1	0.15	766.4631	9	896.7000	3	0.28	2687.0782	
795.5053	1	0.38	794.4960	10	906.7161	3	0.17	2717.1265	44*
805.3162	1	0.56	804.3089	11*	979.3175	3	0.45	2934.9307	45*
810.5579	1	0.18	809.5506	12	986.6416	33	0.08	2956.9030	Adduct Na ⁺ of 45
818.4157	1	0.07	817.4084	13	1014.1115	3	0.11	3039.3127	46*
823.3649	1	0.08	822.3576	14	1033.7432	3	0.07	3098.2078	47
846.4436	1	0.42	845.4363	15	1039.0695	3	0.58	3114.1867	48*
851.3978	1	0.33	850.3905	16	1170.1041	3	7.32	3507.2905	49*
854.5838	1	0.12	853.5765	17	877.8348	4	0.38	3507.3101	
868.3716	1	0.38	867.3643	18	1175.4335	3	0.28	3523.2787	50
998.3748	1	0.08	997.3675	19*	1177.4319	3	0.46	3529.2739	51*
1006.3535	1	0.30	1005.3462	20*	883.3284	4	0.12	3529.2845	
507.6903	2	0.34	1013.3660	21*	1180.7644	3	0.19	3539.2714	52
515.6881	2	0.15	1029.3616	22*	887.3198	4	0.09	3545.2501	53*
543.2088	2	0.38	1084.4030	23*	1184.1069	3	0.29	3549.2989	54*
563.8521	2	0.85	1125.6896	24*	888.3371	4	0.17	3549.3193	
624.2387	2	0.11	1246.4628	25	1191.4551	3	0.51	3571.3435	
645.7581	2	0.08	1289.5016	26	893.8458	4	1.62	3571.3541	55*
646.7568	2	0.18	1291.4990	27	715.2792	5	0.30	3571.3596	
1328.5372	1	0.19	1327.5299	28*	1242.5383	3	1.22	3724.5931	
664.7733	2	2.42	1327.5320		932.1603	4	6.50	3724.6121	56*
1329.5195	1	1.85	1328.5122	29*	745.9301	5	0.82	3724.6141	
665.2647	2	18.98	1328.5148		936.1586	4	0.38	3740.6053	Oxidation of 56*
676.2563	2	2.53	1350.4980	Adduct Na ⁺ of 29	749.1303	5	0.13	3740.6151	
1359.5611	1	2.66	1358.5538	30*	937.6563	4	0.17	3746.5961	57
680.2878	2	28.22	1358.5610		1253.8698	3	0.33	3758.5876	
684.2411	2	0.29	1366.4676	Adduct K ⁺ of 29*	940.6553	4	1.80	3758.5921	58*
687.2608	2	1.01	1372.5070	Adduct 2Na ⁺ of 29*	752.7267	5	0.61	3758.5971	
691.2791	2	4.21	1380.5436	Adduct Na ⁺ of 30*	944.6572	4	0.14	3774.5997	Oxidation of 58*
699.2645	2	0.47	1396.5144	Adduct K ⁺ of 30*	1308.1698	3	0.17	3921.4876	59
702.2843	2	0.52	1402.5540	Adduct 2 Na ⁺ of 30	987.8723	4	0.12	3947.4601	60*
706.7946	2	0.20	1411.5746	31*	991.8697	4	0.14	3963.4497	Oxidation of 60*
739.2960	2	0.12	1476.5774	32*	1339.5401	3	0.08	4015.5985	61
764.7840	2	0.05	1527.5534	33*	1349.5451	3	0.06	4045.6135	62
809.2464	2	0.15	1616.4782	34*	1035.7029	4	0.22	4138.7825	63
810.3680	2	0.10	1618.7214	35*	1044.4545	4	0.07	4173.7889	64
540.5819	3	0.20	1618.7239		1135.9669	4	0.15	4539.8385	65

Asterisk (*) indicates the masses, which have also been retrieved in the coupling experiment.

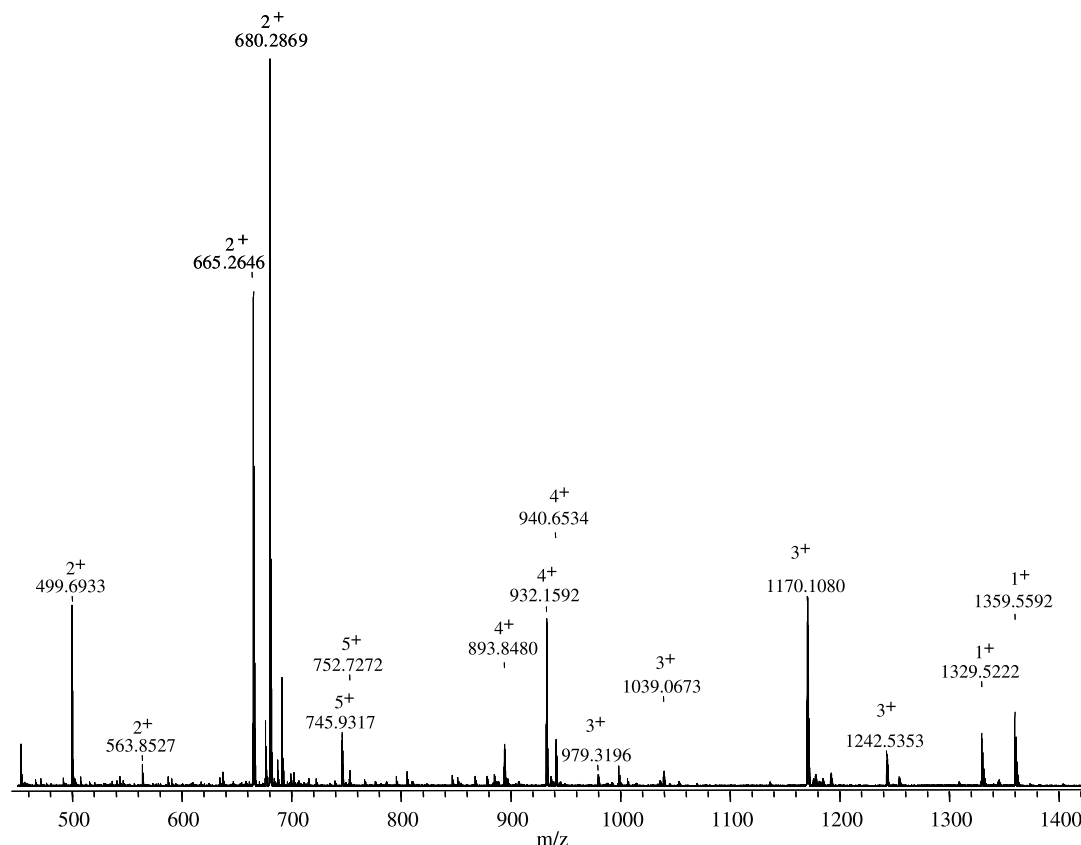


Fig. 1. Nano-ESI-FTICR mass spectrum of *Conus virgo* crude venom.

compounds of interest from these experiments were 1328.5142 and 1358.5592 Da, which are very close to those obtained from the nano-ESI experiments. The experimental error of MALDI-FTICR with internal calibration can be estimated to be less than 1 ppm. This accuracy will be further used to confirm the amino acid sequence deduced from MS/MS data.

Secondly, to gain more insight into the sequences of the compounds of interest, the reduction of the disulfide bonds of all species present in the crude venom was undertaken

using TCEP. The shifts in mass, due to the reduction of S–S bonds into free –SH (+2 Da/disulfide bond) were monitored by MALDI-FTICR (Fig. 2) and allowed the number of disulfide bonds to be calculated for each molecular species (external calibration used).

A mass shift of +4 Da was observed for both compounds of interest to give m/z 1333.553 (MH^+) and 1363.598 which can be rationalized by the presence of two disulfide bonds in each compound. This observation is very helpful in order to determine the family of these conotoxins,

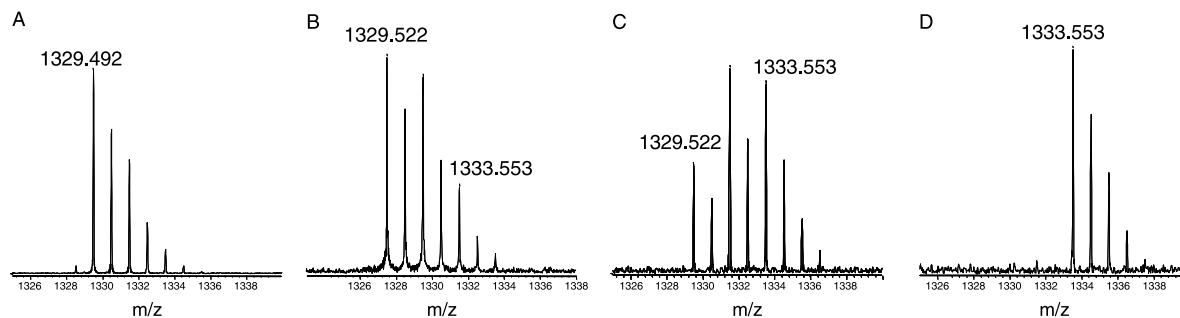


Fig. 2. Disulfide bonds reduction with TCEP of the toxin at m/z 1329.492 followed by MALDI-FTICR. A shift of 4 Da is observed characterizing the reduction of two S–S bonds. (A) $t=0$ min; (B) $t=10$ min; (C) $t=20$ min; (D) $t=30$ min.

since only the α -conotoxins and the T-superfamily, which includes χ -conotoxins and toxins with the disulfide framework V, are known to possess two disulfide bonds in this molecular mass range. From a biological point of view, the α -conotoxins are known to target the nicotinic acetylcholine receptors (Dwoskin and Crooks, 2001) and the χ -conotoxins inhibit the noradrenaline transporter (Sharpe et al., 2001) whereas the peptides belonging to the disulfide framework V are likely to be functionally diverse with strikingly divergent sequences and varying extents of post-translational modifications (Walker et al., 1999).

3.2. De novo sequencing by mass spectrometry

In order to obtain the sequence of the two conotoxins, nano-ESI-MS/MS experiments were carried out on the reduced species. Although the doubly charged ions were the major ones in the oxidized forms of the peptides, only singly charged ions were observed after the reduction step. For that reason, we had to abandon the electron capture dissociation (Cooper et al., 2005) experiments initially planned. These results appear surprising at first, as it is commonly assumed that for proteins, the reduction of S–S bonds facilitates protonation, but this decrease of charge state after disulfide reduction has already been reported (Quinton et al., 2005). Actually, it is difficult to compare proteins and peptides and it is not clear whether for small molecules such as peptides,

the reduction of S–S bonds yields a really different structure in which new sites of protonation become accessible.

Therefore, the MS/MS experiments were undertaken on the singly charged ions at m/z 1333.5497 and 1363.6038 (MH^+) directly from the crude venom. These spectra are reported in Fig. 3A and B, respectively.

3.3. The sequences share similarities

The first observation is that the two tandem mass spectra share similarities. The 30 Da difference between parent ions (m/z 1333.5497 and 1363.6038) can be observed also between some fragment ions, indicating a partially common structure for both toxins. Moreover, the comparison between both spectra also shows that some fragment ions are shifted by 16 or 14 Da ($16 + 14 = 30$ Da) which may imply that only two amino acids are modified between both structures. The search of sequence tags by the de novo software (Bruker) using a low tolerance (0.02 Da) gives poor results since only two tags, relying on the most abundant fragment ions in the spectra, are retrieved for each sequence. (i) A Cys-Cys tag can be retrieved in the two spectra. Note that this tag is often observed for conotoxins since this sequence is present in the O-, M-, A-, T- and I-superfamilies. (ii) The second tag found automatically is different for the two toxins. For the ion at m/z 1333.5497, a Leu/Ile-Thr-Pro-Cys tag is proposed whereas the sequence proposed for m/z 1363.6038 is Leu/Ile-Thr-Leu/Ile-Cys. It is noteworthy that the only

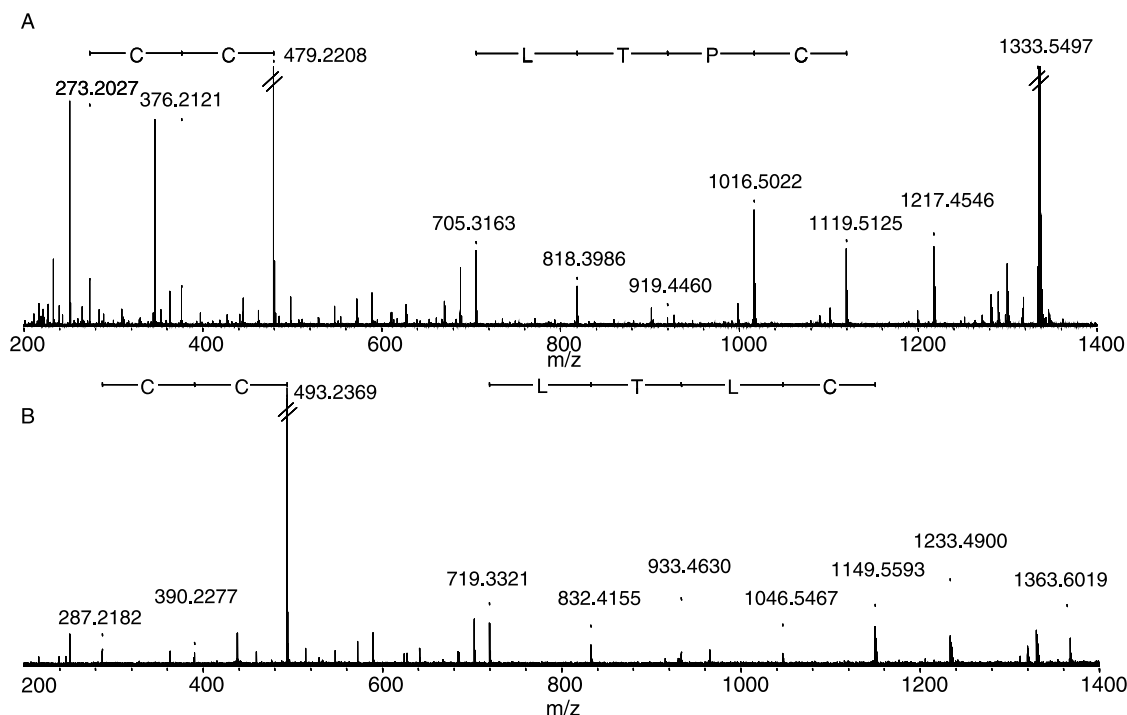


Fig. 3. Tandem mass spectra of the reduced peptides at m/z 1333.5497 (A) and 1363.6038 (B). On top are depicted the results of de novo sequencing by the Bruker software.

difference between these two tags, containing five amino acids, is the presence of a proline in one case and a leucine (or isoleucine) in the other case. The mass difference between these two amino acids is 16 Da, a difference already observed in the comparison of the spectra. Thus, we can suppose that the first difference in both sequences under investigation is this substitution Pro/Leu.

In the MS/MS spectrum of m/z 1363.6038, the fragment ion at m/z 1233.4900 (-130.1138 Da) is accompanied by another ion at m/z 1205.4409 which corresponds to a CO loss, indicative of a *b* ion (Roepstorff and Fohlman, 1984). Thus, the loss of 130.1138 Da can be rationalized if an amidated Leu/Ile is present at the C-terminus (amidation being a common post-translational modification in conotoxins). Mass accuracy is helpful to distinguish between Ile/Leu (which would give a theoretical loss of 130.1106 Da) and hydroxyproline (which would give a theoretical loss of 130.0742 Da). The hydroxyproline yields an error which is not acceptable for FTMS.

A similar fragmentation of the C-terminus is observed for the MS/MS spectrum of m/z 1333.5497, which yields m/z 1217.4546. This 116.0951 Da loss can be rationalized if an amidated valine is the last amino acid of the sequence. The difference between Ile/Leu and Val is exactly 14 Da and could be the second modification between the two sequences. Our attempt to find other *b* ions in both spectra failed, and thus, we supposed that all other ions were *y*-type ions. This was confirmed by the observation of satellite ions 17 Da lower in mass for many peaks. At this stage in the de novo analysis, it appeared that the absence of some fragment ions in the spectra precluded determination of the full sequence of both peptides. Therefore, we decided to use the high accuracy of FT-ICR mass spectrometry (about 10 ppm for this kind of experiment) to determine the amino acid composition of all 'gaps' observed in the spectra, taking into

account the various post-translational modifications often encountered for conotoxins.

3.4. Results for m/z 1363.6038

Table 2 shows the comparison of the experimental masses deduced from the MS/MS spectrum of m/z 1363.6038 ion with the theoretical masses of potential tags.

The first 'gap' found is between m/z 493.2369 and 719.3321 which is 226.0952 Da. The only possible tag in this case, is Glu-Pro (or Pro-Glu) with a mass of 226.0954 Da, corresponding to an error of 0.7 ppm. As seen in the table, all other propositions are out of range.

To find the sequence of the other parts of the molecule, we tested several variations of the extremities: free C-terminus, free N-terminus or amidated C-terminus. To explain the m/z 287.2182 ion, the only reasonable possibility is a Arg-Leu(NH₂) tag as the amidated C-terminus with a calculated error of 2.9 ppm. A Leu-Arg(NH₂) tag can be discarded on the basis of the information obtained previously.

Finally, the loss of 214.0426 Da observed between the parent ion and m/z 1149.5593 corresponds to an N-terminal fragmentation. This mass must take into account the presence of a cysteine residue since only three cysteines were attributed in the whole sequence whereas two disulfides bridges were found. The only possible explanation is a pyroglutamic acid at the N-terminus, which yields an error of 6.4 ppm.

Thus, at this point in the study, the MS/MS data allowed us to propose the following primary structure for the first toxin: Pyr-Cys-Cys-Leu/Ile-Thr-Leu/Ile-(Glu + Pro)-Cys-Cys-Arg-Leu/Ile(NH₂). The Cys-Cys-(Xxx)_n-Cys-Cys pattern indicates that this toxin belongs to the T-superfamily

Table 2

Tag identification based on the mass accuracy of the fragment ions (<10 ppm) observed in MS/MS spectrum of m/z 1363.6038

Measured mass	Location within the sequence	Searched mass	Tag	Theoretical mass	Error (ppm)
226.0952	Within	226.0952	LL	226.1681	322.4
			OL	226.1518	250.1
			EP	226.0954	0.7
287.2182	Acidic C-ter	268.2004	HM	268.0994	376.6
			GGGP or GNP	268.1171	310.4
	N-ter	286.2109	GTK or ASK	286.1641	163.5
			AGGT or AAGS	286.1277	290.8
	Amidated C-ter	269.1844	LR	269.1852	2.9
			APT	269.1376	174.0
			AVV or GVL	269.1739	38.8
214.0426	pyroGlu-N-ter	214.0426	OR	269.1488	132.2
			CZ	214.0412	6.4

The searched masses take into account the N-terminus or C-terminus of the sequence [$-OH_2$ for acidic C-ter, $-H$ for N-ter, $-NH_3$ for amidated C-ter, $(-H + H)$ for pyroglutamylated N-ter]. O, hydroxyproline; Z, pyroglutamic acid.

with the cysteine framework V. This toxin was named ViVA.

3.5. Results for m/z 1333.5497

The same process using the high mass accuracy of the instrument was applied to the ion at m/z 1333.5497 (data not shown) and its primary structure was subsequently determined as Pyr-Cys-Cys-Pro-Thr-Leu-Ile-(Glu+Pro)-Cys-Cys-Arg-Val(NH₂) and the toxin named ViVB.

3.6. Glu-Pro or Pro-Glu

For both peptides, the precise location of Glu and Pro within the structure cannot be determined directly on the basis of the MS/MS data. Nevertheless, the high intensity of m/z 479.2208 and 493.2369, corresponding to the cleavage before the second Cys-Cys tag (y_4), can be better rationalized if a Pro-Glu-Cys-Cys chain is present in the sequence rather than Glu-Pro-Cys-Cys for two different reasons. The first one is the well known enhancement of fragmentation at the C-terminal side of acidic amino acids, especially for singly charged ions where the charge is localized on a basic residue, an arginine in our case (Wysocki et al., 2000). The second reason relies on the fact that favored fragmentation on N-terminus of prolines is often observed (Breci

et al., 2003). This would explain the lack of fragmentation between Pro and Glu (y_5). To confirm this hypothesis, the two possible peptides having the following structures: ZCCPTLPECCRV(NH₂) and ZCCPTLEPCCRV(NH₂) (Z=Pyr) were synthesized and their MS/MS spectra were acquired in the same conditions than for the native toxins. The resulting spectra are shown in Fig. 4A and B.

The two spectra exhibit similar fragmentation patterns, due to a partially common sequence, but two major differences can be highlighted. First, for the sequence with Glu-Pro-Cys-Cys (Fig. 4A), the y_5 ion is the most abundant whereas this fragmentation is not observed for the other structure (Fig. 4B). The second notable difference is the presence of an intense y_4 ion in the second spectrum, which is very low in the first one. The comparison between these spectra and the spectrum depicted in Fig. 3B indicates without ambiguity that the native peptide contains a Pro-Glu-Cys-Cys tag. The presence of an intense y_4 ion and the absence of y_5 in the MS/MS spectrum obtained for the other native peptide (Fig. 3A) allow us to draw the same conclusion for both conotoxins. Table 3 summarizes these results.

It is noteworthy that the theoretical monoisotopic masses calculated from the elemental composition of both compounds correspond to the experimental ones within a 1 ppm error as expected. The presence of an

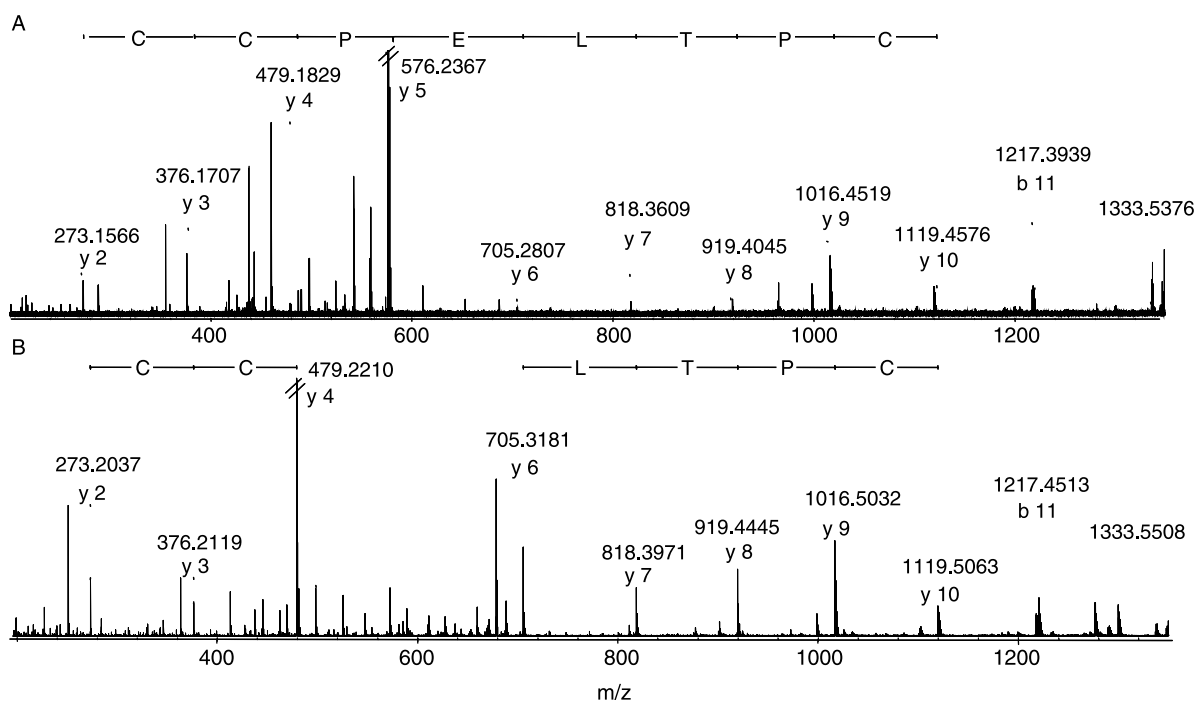


Fig. 4. MS/MS spectra of the reduced synthetic peptides (A) ZCCPELTTPCCRV(NH₂) and (B) ZCCPELTPCCRV(NH₂) (Z=pyroglutamic acid).

Table 3

Sequence of both peptides with theoretical monoisotopic molecular masses ($M_{r_{\text{theo}}}$) and experimental monoisotopic molecular masses ($M_{r_{\text{exp}}}$) in Dalton

Peptide sequence	ViVA [Z-C-C-I/L-T-I/L-P-E-C-C-R-I/L(NH ₂)]	ViVB [Z-C-C-P-T-I/L-P-E-C-C-R-V(NH ₂)]
$M_{r_{\text{theo}}}$ (Da)	1358.5604	1328.5134
$M_{r_{\text{exp}}}$ (Da)	1358.5592	1328.5142
Error (ppm)	0.9	0.6
Fragment ions observed		

Error (in ppm) corresponds to the difference between $M_{r_{\text{theo}}}$ and $M_{r_{\text{exp}}}$. Z, pyroglutamic acid.

Arg as the antepenultimate amino acid in the sequence explains the predominance of y ions and the presence of only b_{11} .

3.7. LC/MS results

Our results show that a direct approach by nano-ESI-FTMS on the crude venom allows both peptide mass fingerprinting and de novo sequencing. Nevertheless, the wide mass range of the different species present in the venom and the fact that they can have various physicochemical properties emphasizes the problem of ion suppression due to the use of nano-electrospray ionization. Although the use of nano-electrospray, instead of electrospray, reduces this phenomenon by increasing the ionization efficiency (Wilm and Mann, 1996; Juraschek et al., 1999), we decided to compare our mass fingerprinting results with those obtained with a classical LC/MS approach. For that purpose, we performed nano-LC-FTMS experiments. The reconstructed base peak chromatogram obtained (with classical experimental conditions for this kind of analysis) is shown in Fig. 5 together with the extracted ion chromatograms corresponding to the two conotoxins of interest.

First of all, the analysis of this chromatogram allows the determination of about 130 different molecular masses in the range 400–7000 Da which is much more than in the direct nano-ESI-FTMS approach. Nevertheless, the most abundant compounds observed in the direct approach are still the major ones with the nano-LC coupling. This is illustrated by the extracted ion chromatograms corresponding to $MH^+1329.5$ and 1359.6.

Among the 130 molecular masses, 45 are common with the direct approach. The intensity of these 45 peaks represents 90% of the total signal in the direct experiment and 71% in LC/MS, which shows that the new masses measured correspond to compounds of low abundance.

A more striking observation is that 21 compounds observed by nano-ESI-FTMS are not found in the coupling experiments. This phenomenon has been observed previously by Taggi et al. (2004) who showed that the experimental conditions for the chromatographic separation are of prime importance to recover all the compounds present in a crude venom. This suggests that using directly mass spectrometry (or other spectroscopic techniques) on crude extracts, as we have done, avoids some discrimination against some classes of compounds and can be complementary to the coupling. A closer examination of the different masses lost between both kinds of experiments indicates that the discrimination observed cannot be correlated to a specific molecular mass range. The results, gathered in Table 4, give two different informations: (i) the coupling allows new masses to be measured for low molecular compounds as well as for heavier ones; (ii) some compounds are lost during the chromatographic step, regardless of their molecular weight. To gain more insight into this problem, we compared the results obtained for nano-LC/MS and direct nano-ESI experiments for the tryptic digest of a known protein. The experimental conditions were the same as those used for this study. Similar results are obtained since many tryptic peptides (of low and high molecular masses) observed in the spectrum obtained from direct infusion are not retrieved by using the coupling. Thus, we can suppose that the chromatographic step is responsible for the loss of material observed. Further investigations are on the way to evaluate the role of the precolumn (solvent used for the injection, rinsing duration, flow) on this phenomenon.

4. Conclusion

In this work, the crude venom of *Conus virgo* was analyzed by Fourier transform mass spectrometry (FTMS). Nano-electrospray analysis of the crude venom, without chromatographic separation, allowed about 64 distinct molecular masses in the range 500–4500 Da to be measured.

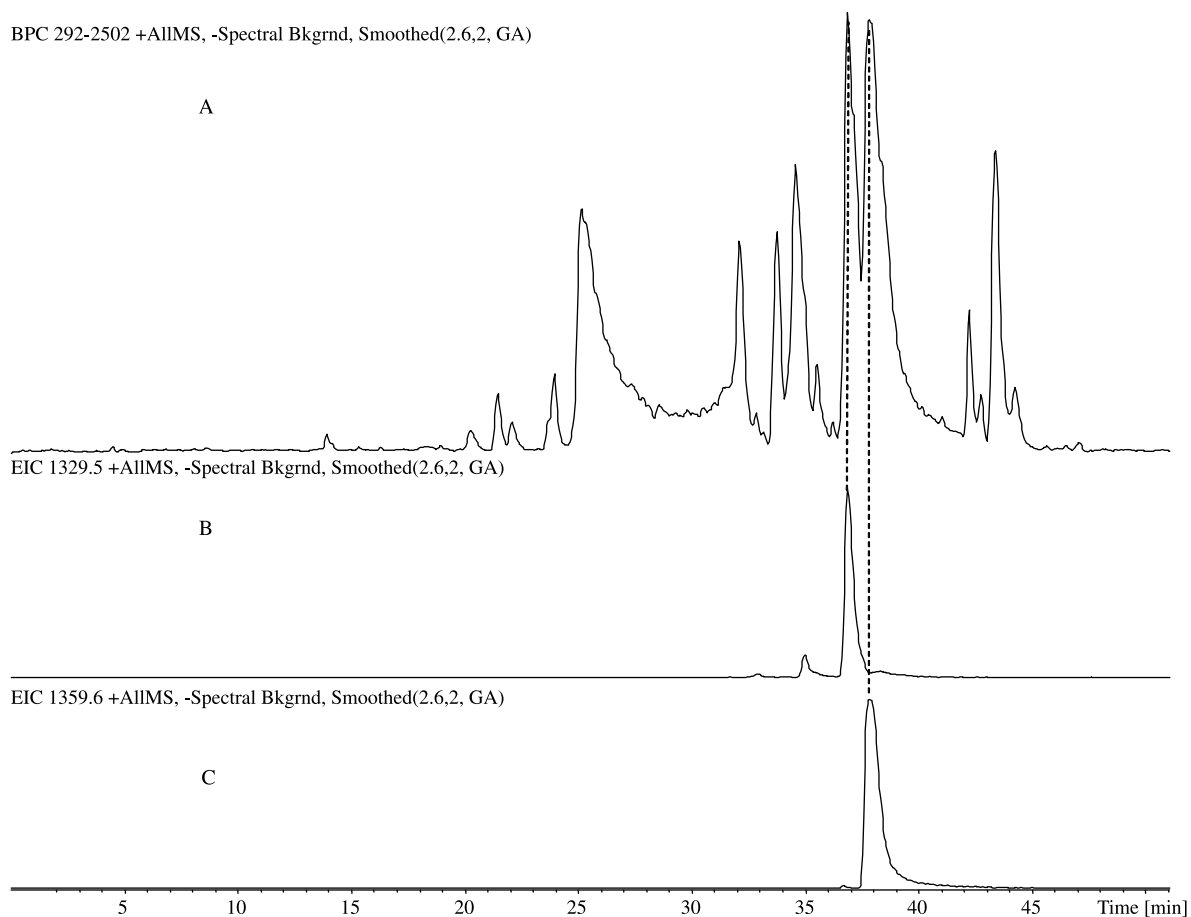
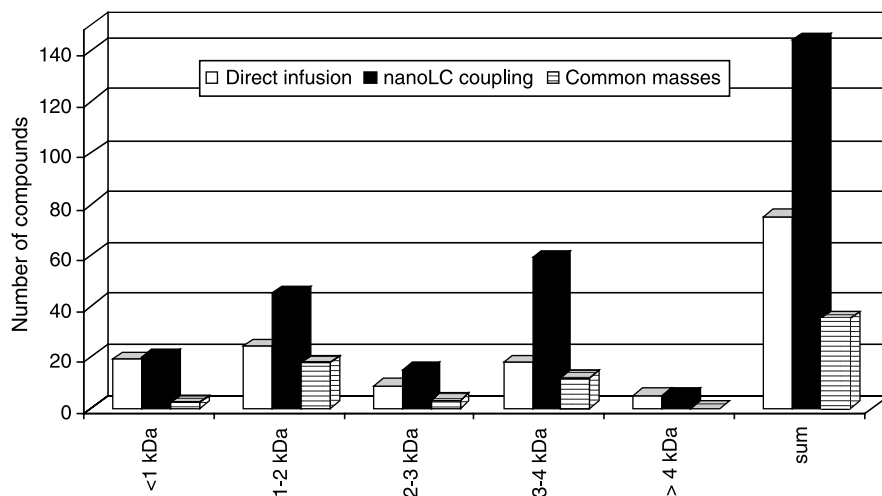


Fig. 5. (A) Reconstructed base peak chromatogram obtained for the venom of *Conus virgo* by nano-LC-FTMS; (B) extracted ion chromatogram for m/z 1329.5; (C) extracted ion chromatogram for m/z 1359.6.

Table 4

Comparison between results obtained for direct infusion and nano-LC coupling for the crude venom in term of number of compounds observed related to different mass ranges



All species detected were in a large dynamic range and two compounds were predominant. MALDI experiments with internal calibration were then used to determine the molecular masses of these two components with the highest mass accuracy possible (<1 ppm): 1328.5142 and 1358.5592 Da, named, respectively, ViVA and ViVB. Then, the whole venom was reduced and both compounds of interest were de novo sequenced within the complex mixture. It is worth to note that the mass accuracy and resolution provided by FTMS were necessary to complete the sequencing, which gave the following results:

ViVa: Z–C–C–I/L–T–I/L–P–E–C–C–R–I/L(NH₂)

ViVB: Z–C–C–P–T–I/L–P–E–C–C–R–V(NH₂)

The disulfide bond pattern Cys-Cys-(Xxx)₅-Cys-Cys clearly identified these two peptides as conotoxins belonging to the T-superfamily, framework V. For that reason, we can suppose that the connectivity between the Cys residues is 1–3 and 2–4, but this should be confirmed, as well as the biological activity of both peptides. Unfortunately the discrimination Ile/Leu was not possible on the basis of our results and high-energy mass analyses are being conducted to permit complete sequencing. As often observed for this class of compound, both conotoxins are highly post-translationally modified: amidated C-terminus, pyroglutamic acid residue at the N-terminus and two disulfide bonds. Our experiments indicate that ViVA and ViVB are the most abundant toxins in *Conus virgo*'s venom. This implies that the biological activity of these two toxins is probably very efficient, at least on worms, allowing this cone snail to capture its preys.

Complementary nano-LC/FTMS experiments were undertaken and the results were compared to the direct approach by nano-ESI. Among the 130 molecular masses found, only 45 (the major ones) are common with the direct approach, which means that 21 compounds observed by nano-ESI-FTMS are not found in the coupling experiments. This clearly shows that some discriminations against some classes of compounds occur when a chromatographic step is used before mass spectrometry. In that context, the use of high-resolution mass spectrometry such as FTICR for venom analysis constitutes a fast, straightforward and attractive approach for discovering new toxins within crude venoms.

Acknowledgements

We want to thank Dr Gilles Mourier (CEA, Saclay) and Dr Paulo Marcelo (ESPCI, Paris) for their assistance and helpful discussions.

References

- Bao, Y., Bu, P., Jin, L., Hongxia, W., Yang, Q., An, L., 2005. Purification, characterization and gene cloning of a novel phospholipase A₂ from the venom of *Agkistrodon blomhoffii ussurensis*. *Int. J. Biochem. Cell Biol.* 37, 558–565.
- Breci, L.A., Tabb, D.L., Yates III, J.R., Wysocki, V.H., 2003. Cleavage N-terminal to proline: analysis of a database of peptide tandem mass spectra. *Anal. Chem.* 75, 1963–1971.
- Burmeister Getz, E.L., Xiao, M., Chakrabarty, T., Cooke, R., Selvin, P.R., 1999. A comparison between the sulfhydryl reductants tris(2-carboxyethyl)phosphine and dithiothreitol for use in protein biochemistry. *Anal. Chem.* 71, 73–80.
- Burn, J.A., Butler, J.C., Moran, J., Whitesides, G.M., 1991. Selective reduction of disulfides by tris(2-carboxyethyl)phosphine. *J. Org. Chem.* 56, 2648–2650.
- Caravatti, P., Allemann, M., 1991. The infinity cell—a new trapped-ion cell with radiofrequency covered trapping electrodes for fourier-transform ion-cyclotron resonance mass-spectrometry. *Org. Mass Spectrom.* 26, 514–518.
- Cooper, H.J., Hakansson, K., Marshall, A.G., 2005. The role of electron capture dissociation in biomolecular analysis. *Mass Spectrom. Rev.* 24, 201–222.
- Diego-García, E., Batista, C.V.F., García-Gómez, B.I., Lucas, S., Candido, D.M., Gómez-Lagunas, F., Possani, L.D., 2005. The Brazilian scorpion *Tityus costatus* Karsch: genes, peptides and function. *Toxicon* 45, 273–283.
- Dwoskin, L.P., Crooks, P.A., 2001. Competitive neuronal nicotinic receptor antagonists: a new direction for drug discovery. *J. Pharmacol. Exp. Ther.* 298, 395–402.
- Gauthier, J.W., Trautman, T.R., Jacobson, D.B., 1991. Sustained off-resonance irradiation for collision-activated dissociation involving fourier transform mass spectrometry. Collision-activated dissociation technique that emulates infrared multiphoton dissociation. *Anal. Chem. Acta* 246, 211–225.
- Gorman, J.J., Wallis, T.P., Pitt, J.J., 2002. Protein disulfide bond determination by mass spectrometry. *Mass Spectrom. Rev.* 21, 183–216.
- Hansson, K., Furie, B., Furie, B.C., Stenflo, J., 2004. Isolation and characterization of three novel Gla-containing conus marmoreus venom peptides, one with a novel cysteine pattern. *Biochem. Biophys. Res. Commun.* 319, 1081–1087.
- Juraschek, R., Dulcks, T., Karas, M., 1999. Nanoelectrospray—more than just a minimized-flow electrospray ionization source. *J. Am. Soc. Mass Spectrom.* 10, 300–308.
- Kalume, D.E., Stenflo, J., Czerwiec, E., Hambe, B., Furie, B.C., Furie, B., Roepstorff, P., 2000. Structure determination of two conotoxins from conus textile by a combination of matrix-assisted laser desorption/ionization time-of-flight and electrospray mass spectrometry and biochemical methods. *J. Mass Spectrom.* 35, 145–156.
- Kelleher, N.L., Zubarev, R.A., Bush, K., Furie, B.C., Mc Lafferty, F.W., Walsh, C.T., 1999. Localization of labile posttranslational modifications by electron capture dissociation: the case of gamma-carboxyglutamic acid. *Anal. Chem.* 71, 4250–4253.
- Li, L., Masselon, C.D., Anderson, G.A., Pasa-Tolic, L., Lee, S.-W., Shen, Y., Zhao, R., Lipton, M.S., Conrads, T.P., Tolic, N., Smith, R.D., 2001. High-throughput peptide identification from protein digests using data-dependent

- multiplexed tandem FTICR mass spectrometry coupled with capillary liquid chromatography. *Anal. Chem.* 73, 3312–3322.
- Liska, A.J., Shevchenko, A., 2003. Combining mass spectrometry with database interrogation strategies in proteomics. *Trends Anal. Chem.* 22, 291–298.
- Marshall, A.G., 2004. Accurate mass measurement: taking full advantage of nature's isotopic complexity. *Physica B* 346–347, 503–508.
- Marshall, A.G., Hendrickson, C.L., 2002. Fourier transform ion cyclotron resonance detection: principles and experimental configurations. *Int. J. Mass Spectrom.* 215, 59–75.
- Marshall, A.G., Hendrickson, C.L., Jackson, G.S., 1998. Fourier transform ion cyclotron resonance mass spectrometry: a primer. *Mass Spectrom. Rev.* 17, 1–35.
- Olivera, B.M., 1997. Conus venom peptides, receptor and ion channel targets, and drug design: 50 million years of neuropharmacology. *Mol. Biol. Cell* 8, 2101–2109.
- Olivera, B.M., Rivier, J., Clark, C., Ramilo, C.A., Corpuz, G.P., Abogadie, F.C., Mena, E.E., Woodward, S.R., Hillyard, D.R., Cruz, L.J., 1990. Diversity of *Conus* neuropeptides. *Science* 249, 257–263.
- Quinton, L., Le Caer, J.-P., Phan, G., Ligny-Lemaire, C., Bourdais-Jomaron, J., Ducancel, F., Chamot-Rooke, J., 2005. Characterization of new toxins within crude venoms by combined use of fourier transform mass spectrometry and cloning. *Anal. Chem.* 77, 6630–6639.
- Roepstorff, P., Fohlman, J., 1984. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed. Mass Spectrom.* 11, 601.
- Schlosser, A., Lehmann, W.D., 2002. Patchwork peptide sequencing: extraction of sequence information from accurate mass data of peptide tandem mass spectra recorded at high resolution. *Proteomics* 2, 524–533.
- Sharpe, I.A., Gehrmann, J., Loughnan, M.L., Thomas, L., Adams, D.A., Atkins, A., Palant, E., Craik, D.J., Adams, D.J., Alewood, P.F., Lewis, R.J., 2001. Two new classes of conopeptides inhibit the α -1-adrenoreceptor and noradrenaline transporter. *Nat. Neurosci.* 4, 902–907.
- Spengler, B., 2004. De novo sequencing, peptide composition analysis, and composition-based sequencing: a new strategy employing accurate mass determination by fourier transform ion cyclotron resonance mass spectrometry. *J. Am. Soc. Mass Spectrom.* 15, 704–715.
- Taggi, A.E., Meinwald, J., Schroeder, F.C., 2004. A new approach to natural products discovery exemplified by the identification of sulfated nucleosides in spider venom. *J. Am. Chem. Soc.* 126, 10364–10369.
- Tzouros, M., Manov, N., Bienz, S., Bigler, L., 2004. Tandem mass spectrometric investigation of acylpolyamines of spider venoms and their ^{15}N -labeled derivatives. *J. Am. Soc. Mass Spectrom.* 11, 1636–1643.
- Walker, C.S., Steel, D., Jacobsen, R.B., Lirazan, M.B., Cruz, L.J., Hooper, D., Shetty, R., Delacruz, R.C., Nielsen, J.S., Zhou, L.M., Bandyopadhyay, P., Craig, A.G., Olivera, B.M., 1999. The T-superfamily of conotoxins. *J. Biol. Chem.* 274, 30664–30671.
- Wermelinger, L.S., Dutra, D.L.S., Oliveira-Carvalho, A.L., Soraes, M.R., Bloch Jr., C., Zingali, R.B., 2005. Fast analysis of low molecular mass compounds present in snake venoms: identification of ten new pyroglutamate-containing peptides. *Rapid Commun. Mass Spectrom.* 19, 1703–1708.
- Wilm, M., Mann, M., 1996. Analytical properties of the nanoelectrospray source. *Anal. Chem.* 68, 1–9.
- Wysocki, V.H., Tsaprailis, G., Smith, L.L., Brei, L.A., 2000. Mobile and localized protons: a framework for understanding peptide dissociation. *J. Mass Spectrom.* 35, 1399–1406.
- Yates III, J.R., 2000. Mass spectrometry: from genomics to proteomics. *Trends Genet.* 16, 5–8.